Essays in behavioural economics

Gneezy, U.

Publication date:
1997

Document Version
Publisher's PDF, also known as Version of record

CentER
for
Economic Research

Essays in Behavioral Economics

Uri Gneezy

Tilburg University

Essays in Behavioral Economics

Uri Gneezy

Essays in Behavioral Economics

Proefschrift ter verkrijging van graad van doctor aan de Katholieke Universiteit Braban, op gezag van de rector magnificus, prof. dr. L. F. W. de Klerk, in het openbaar te verdedigen ten overstaan van een door het college van decanen aangewezen commissie in de aula van de Universiteit op vrijdag 17 oktober 1997 om 16.15 uur door

Uri Gneezy

geboren op 6 juni 1967 te Tel-Aviv, Israel.

Promotor: Prof. dr. E.E.C. van Damme

לאיילת

Preface

Background

Neoclassical economics is based on a model of a rational decision maker who maximises his utility. However, a growing body of empirical evidence show that the rational decision making model fails to describe how real people behave. The question economists face is whether the empirical facts should be allowed to spoil the good story. I think they should.

A relatively new area of research in economics, which can be called "behavioral economics", is aimed at closing that gap by improving the descriptive power of models. While normative models are typically based on a set of "rational" assumptions, the descriptive models are based on assumptions which are motivated by observed behavior.

The first step in building better descriptive models is finding the relevant behavioral regularities. For that purpose economists adopted a tool, which psychologists have been using for a very long time: *Experiments*.

This thesis contains a collection of papers which form my first steps into this world.

Acknowledgements

I have spent the last four years at the CentER for Economic Research, learning about science. I wish to thank my advisor, Eric van Damme, for guiding me into this world, and Arie Kapteyn for his help in setting the direction.

Working so closely with people is one of the more enjoyable parts of doing science. I had the pleasure to work on different chapters of this thesis with Jan Potters, Marcel Das, Chaim Fershtman, Werner Guth, and Martin Dufwenberg. I have benefited a lot from discussions with Andreas Blume, Jan Bouckaert, Gilbert van Hagen, Jos Jansen, Aldo Rustichini, Frank Verboven and Peter Wakker. I am also grateful to the members of my Ph.D. committee--Eric van Damme, Chaim Fershtman, Werner Guth, Aldo Rustichini, Arthur Schram, Reinhard Selten, and Peter Wakker-- for their time and effort.

I will always be in debt for the hospitality of CentER, where I met all these people and many others. CentER's staff, apart from creating the pleasant environment, facilitates our work, and enabled us to concentrate on doing research.

Finally, I wish to thank my parents, who made me curious of the world, and my wife, who not only participated very actively in my research, but also, together with our two daughters, made life fun.

Contents

Part I: Behavioral Finance

Part II: Behavioral Game Theory

Chapter 1

An Experiment on Risk Taking and Evaluation Periods[1]

Abstract: Does the period over which individuals evaluate outcomes influence their investment in risky assets? Results from this study show that the more frequently returns are evaluated, the more risk averse investors will be. The results are in line with the behavioral hypothesis of "myopic loss aversion," which assumes that people are myopic in evaluating outcomes over time, and are more sensitive to losses than to gains. The results have relevance for the equity premium puzzle, and also for the marketing strategies of fund managers.

---

[1] This is a joint project with Jan Potters. The paper was first published in the *Quarterly Journal of Economics*, May 1997.

1. Introduction

Recently, Benartzi and Thaler (1995) put forward an explanation for the equity premium puzzle. This puzzle refers to the fact that over the last century the risk-return relationship has been so much more favorable for stocks than for bonds, that unreasonably high levels of risk aversion would be needed to explain why investors are willing to hold bonds at all (Mehra and Prescott, 1985). The explanation for this puzzle, advanced by Benartzi and Thaler, is called myopic loss aversion (MLA), and rests on the combination of two behavioral concepts. The first concept is *loss aversion* (Kahneman and Tversky, 1979; Tversky and Kahneman, 1992), which refers to the tendency of individuals to weigh losses more heavily than gains. The second concept is *mental accounting* (Kahneman and Tversky, 1984; Thaler, 1985), which refers to the implicit methods people employ to code and evaluate financial outcomes.

The effect of combining these two concepts is perhaps best illustrated by means of a well-known problem devised by Samuelson (1963). He asked a colleague whether he would be willing to accept a gamble in which there are equal chances to win $200 and to lose $100. The colleague declined this single gamble, but at the same time expressed a willingness to accept multiple plays of the gamble. Although such a preference may have much intuitive appeal, Samuelson proved a theorem, saying that if the single gamble is rejected at every relevant wealth position, then accepting the multiple gamble is inconsistent with expected utility maximization (see Tversky and Bar-Hillel, 1983, for further discussion).

Benartzi and Thaler show that rejecting each single gamble, but accepting a sequence of such gambles is consistent with MLA (see Kahneman and Lovallo, 1993, for a similar argument). If returns are evaluated over a longer period of time, multiple gambles become more attractive due to the lower probability that a loss will be experienced. To illustrate, suppose that the individual is characterized by loss aversion and has a utility function $u(z) = z$ for $z \geq 0$ and $u(z) = 2.5z$ for $z < 0$, where z is the change in wealth due to the gamble. Then, the expected utility of one gamble is negative: $1/2(200) + 1/2(-250) < 0$. Hence, the individual will reject one gamble, and also two gambles--if each is evaluated separately. The same individual, however, accepts two gambles if (s)he evaluates them in combination: $1/4(400) + 1/2(100) + 1/4(-500) > 0$. Hence, rejecting a single gamble while accepting two gambles is quite easily explained by the combined hypotheses of individuals being more sensitive to losses than to gains and evaluating the outcomes of the sequence of gambles in combination.

As the example illustrates, predicts that the dynamic aggregation rules that people employ influence their attitude toward risk. In particular, the period over which individuals evaluate financial outcomes influences their investments in risky assets. By means of theoretical simulations, Benartzi and Thaler show that MLA could thus provide an explanation for the equity premium puzzle. In particular, they show that the size of the equity premium is consistent with investors evaluating their portfolios annually and weighing losses about twice as heavily as gains.

However, Benartzi and Thaler do not present direct (experimental) evidence for the presence of MIA The evidence presented in Benartzi and Thaler is only circumstantial. Hence, we seem to have a choice anomaly-a choice rule that departs from standard theory-that could potentially explain an important phenomenon. Yet, there are no direct and controlled tests that indicate that the anomaly is real. Designing such a test is the purpose of this paper.[2]

We have experimental subjects making a sequence of risky choices. To analyze the presence of MLA, we do not try to estimate the period over which subjects evaluate financial outcomes, but rather we try to manipulate this evaluation period. In our setup, two groups of participants are subjected to the same sequence of choices. Subjects in the first (high frequency) group are supplied with feedback information after each round of the sequence, and can change their choice after each round. The subjects in the second (low frequency) group, however, get feedback information only after three rounds, and can only adapt their choices after three rounds. If our design is successful in manipulating subjects' evaluation period, MLA would predict that the low-frequency subjects make more risky choices. If subjects use a longer horizon to evaluate outcomes, the trade-off between losses and gains becomes more favorable for the risky option. At the same time, subjective expected utility (SEU) theory does not predict a systematic difference in risk taking between the two treatments in our setup.

The remainder of this paper is organized as follows. The next section explains and motivates the design of the experimental test, and spells out the hypothesis. Section 3 presents the results, and Section 4 concludes.

[2] Independently, Thaler, Tversky, Kahneman, and Schwartz (1997) conducted a similar experiment.

2. Design and procedure

Consider an individual who is confronted with a sequence of three independent but identical lotteries, in which there is a probability of 2/3 to lose \$1 and a probability of 1/3 to win \$2.50. If, as is hypothesized by MLA, the individual weighs losses more heavily than gains, then the attractiveness of the lotteries may depend on whether the financial consequences of the gambles are evaluated separately or in combination. For illustration, suppose that the individual weighs losses relative to gains at a rate of $\kappa > 1$. Then the expected utility of a single lottery is $(2/3)\kappa(-1) + (1/3)(2.5)$, which is positive only if $\kappa < 1.25$. However, if a subject evaluates the three lotteries in combination, then the expected utility is $(1/27)(7.5) + (6/27)(4) + (12/27)(0.5) + (8/27)\kappa(-3)$, which is positive if $\kappa < 1.56$. This is because the probability of a loss decreases from 0.67 for a single lottery, to $(0.67)^3 = 0.30$ for three consecutive lotteries. If the financial consequences of the three lotteries are evaluated in combination rather than separately, then the lotteries should become more attractive.[3] It is this basic prediction of MLA that we tested in our experiment, by manipulating the evaluation period of subjects.

In the experiment, subjects were confronted with a sequence of twelve identical but independent rounds of a lottery (betting game). In each of the first nine rounds ("part 1" of the experiment), subjects were endowed with 200 cents.[4] They had to decide which part (X_t) of this endowment they wanted to bet in the lottery ($0 \leq X_t \leq 200$, $t = 1, \ldots, 9$). In the lottery there was

[3] This prediction only depends on losses weighing more heavfiy than gains, and not on the utility function being piecewise linear.

[4] At the time of the experiment, 1 guilder (100 cents) exchanged for about US \$0.60.

a probability of 2/3 of losing the amount bet and a probability of 1/3 of winning two and a half times the amount bet. Subjects were fully informed about the objective probabilities of winning and losing, and about the corresponding size of gains and losses. It is important to stress that subjects could not bet any money accumulated in previous rounds. Hence, the maximum bet in each round is 200 cents, independently of the outcome of the bet in any of the previous rounds. In rounds 10 through 12 ("part 2" of the experiment) subjects were no longer endowed with any additional money from the experimenters. Rather, they had to make bets from the money earned in part 1. To that purpose, a subject's earnings in the nine rounds of part 1 were first totalled and then divided by three. The resulting amount was a subject's endowment (S) for each of the three rounds of part 2. Again, for each round a subject had to decide which part (X_t) of the endowment S to bet in the lottery ($0 \leq X_t \leq S, t = 10, 11, 12$).

The crucial feature of the design is that there were two different treatments: Treatment H (high frequency) and Treatment L (low frequency). In Treatment H the subjects played the rounds one by one. At the beginning of round 1 they had to choose how much of their endowment of 200 cents to bet in the lottery. Then they were informed about the realization of the lottery in round 1. Only then could they decide how much of their new endowment of 200 cents to bet for round 2, and so on. Hence, in this treatment subjects made nine betting decisions in part 1 and three decisions in part 2. In Treatment L, however, subjects played the rounds in blocks of three. At the beginning of round 1, subjects had to decide how much of their endowment of 200 cents to bet in the lotteries of rounds 1, 2, and 3. In addition, these bets were restricted to be equal. If a subject bet X in round 1, (s)he also bet X in rounds 2 and 3 (that is, X_1

$= X_2 = X_3$, with $0 \leq X_t \leq 200$). After subjects decided on their bets, they were informed about the combined realization for rounds 1, 2, and 3. That is, they could not assign a gain or loss to any particular round, but only knew the aggregate result. Subsequently, subjects decided how much to bet in rounds 4, 5, 6, and so on. Hence, in Treatment L subjects make three decisions in part 1, and one decision in part 2.

In Treatment L subjects chose their bet for the next three rounds; they had, therefore, less freedom because they could not change their decision after every round. In particular, by the design of Treatment L we have $X_t = X_{t+1} = X_{t+2}$ for t = 1, 4, 7, 10. In Treatment H these equalities need not hold. Furthermore, the subjects in Treatment H were supplied with more information than were the subjects in Treatment L. When deciding on X_t, a subject in Treatment H was always fully informed about the realizations and corresponding earnings of the previous rounds. A subject in Treatment L, however, simultaneously decided about X_t, X_{t+1}, and X_{t+2} (t = 1, 4, 7, 10). A subject had to decide about X_{t+1} (X_{t+2}) without knowing the realization for round t (rounds t and t + 1). Hence, subjects in Treatment L were supplied with less freedom and less information than those in Treatment H.

The basic idea behind the two treatments of our design is to manipulate the evaluation period. In Treatment L the frequency of choice and information feedback was lower than in Treatment H. As a result, the subjects in Treatment L can be expected to evaluate the financial consequences of betting in a more aggregated way. If the subjects are characterized by MLA, this should make them more apt to bet money in the lotteries.[5]

According to subjective expected utility (SEU) theory, there may be a difference between a subject's behavior in Treatment H and Treatment L. This is so because, within each block of three rounds, a subject in Treatment H has more information about the current wealth level in the second and third round than does a subject in Treatment L. A subject in Treatment H observes gains and losses along the way, and, contrary to a subject in Treatment H, can adjust bets accordingly. The effect of this additional information cannot be unambiguously signed in general.[6] In view of the small stakes involved in the experiment, however, the effect is likely to be small indeed. Hence, with the assumption that wealth effects within each block of three rounds on subjects' risk aversion are negligible, the natural hypothesis under SEU is that there are no systematic differences between average bets of the subjects in Treatment H and Treatment L. This we use as our null hypothesis.

Procedure

We had fourteen experimental sessions, seven for each of the two treatments. The experiment was administrated by pen and paper, and held in a seminar room with subjects seated

[5] In principle, it would be possible to draw conclusions from only part 1 of the experiment. However, since the subjects receive the 200 cents endowment from us, it is possible that they do not experience a lost bet as a "real" loss. In part 2 of the experiment subjects bet their own money, "earned" in part 1. Therefore, we expected that the impact of loss aversion (if at all) would be amplified in part 2. On the other hand, in part 2 subjects' wealth positions and experiences are more diverse. Hence, in part 2 we may also expect to find larger individual differences.

[6] Gollier, Lindsey, and Zeckhauser (1997) derive sufficient conditions on the utility function for this information effect to have an unambiguous sign. Translated to our setting, their results indicate that constant relative risk aversion less than 1 would induce more risk taking in Treatment H than in Treatment L. Under constant absolute risk aversion there should be no treatment effect.

far apart. Six different subjects participated in each session (84 subjects in total).[7] Students were recruited from Tilburg University. An announcement in the university bulletin solicited participants for a decision-making experiment of about 40 minutes, with a reward that would depend on their decisions, but which was likely to be somewhere between 5 and 35 Dutch guilders. For each session eight subjects were invited; six would participate in the betting games, one would act as an assistant, and one would serve as a spare in case of a no show.

Upon entering the room, a short standard-type introduction was read to the subjects by the experimenter. Subjects were informed that the experiment would consist of three parts, but that they would be informed about the instructions of part 2 only after part 1 was finished. After the introduction, each subject drew an envelope out of a stack. Six envelopes contained numbered registration forms for part 1 of the experiment; one envelope contained a note with 'assistant," and one had an empty note (the latter envelope was removed when only seven subjects showed up). The assistant was told that he would receive a payment equal to the average earnings of the other participants. The subject who drew the empty note was paid f10 for showing up and was asked to leave the room.

Instructions (in Dutch) for part 1 were distributed and read aloud. After that, subjects could examine the instructions for a few additional minutes, and (privately) ask questions.

Subjects were then asked to record their first bets. The lottery was conducted by the assistant. To determine whether a subject gained or lost in a round's lottery, we used private "win

[7] As it turned out, we had one subject who was in the experiment twice. We delete his second set of choices from the data, leaving us with 41 observations in Treatment H.

letters" which were indicated on the Registration Form. For subjects 1 and 2 the win letter was A, for subjects 3 and 4 it was B, and for subjects 5 and 6 the win letter was C. We used different win letters to introduce more variation in the realization of gains and losses within each session. The assistant used a box containing three disks marked, respectively, A, B, and C. After the subjects had recorded their bets for the round, the assistant first showed the contents of the box to the subjects (to show that the box, in fact, contained an A, B, and C); (s)he then shook the box and randomly took one disk out of the box. The letter on the disk was the so-called round letter for the round. If a subject's private win letter matched the round letter, (s)he won in the lottery; if the letters did not match, (s)he lost. Since there were three letters in the box, only one of which matched a subject's win letter, the probability of winning in any round's lottery was 1/3, and the probability of losing was 2/3.

In Treatment L the subjects fixed bets for three rounds, and three lotteries were conducted by the assistant. For that purpose, the assistant used three boxes, each containing three disks labeled A, B, and C. The assistant first showed the contents of each box to the subjects (to show that each box, in fact, contained an A, B, and C); (s)he then shook the boxes and randomly took one disk out of each box. The three disks drawn (one for each round) were then shown simultaneously to the subjects.[8] The letters on the three disks drawn were the round letters for the present three rounds.

[8] The main purpose of our design is to manipulate the evaluation period of the subjects in Treatment L. We wanted them to evaluate three consecutive lotteries in an aggregated way, without experiencing the losses and gains of each separate lottery. Therefore, the outcomes of the three lotteries were shown to them simultaneously. In this way it was not possible for them to attribute a gain or a loss to any particular round in the block of three.

After each round (three rounds in Treatment L), subjects calculated and recorded their own earnings on their registration form. We checked these calculations to make sure that they understood the procedure, and that they did not cheat. Then subjects recorded their bets for the next round (next three rounds in Treatment L).

At the end of the nine rounds, total earnings were calculated, and forms were collected. The experimenter divided these total earnings by three to determine the starting endowment (maximum bet) for each of the three rounds of part 2. This starting endowment (*S*) was indicated on top of the Registration Form for part 2. These forms were distributed together with the instructions for part 2. The instructions were read aloud, and then the three betting rounds for part 2 were held. Again, subjects calculated their own earnings. After it was finished, all subjects were paid.[9] The assistant was paid the average earnings of the other subjects. That concluded the experiment.

3.Results

Analyzing the results of part 1 is a straightforward exercise. We simply compare the average percentage of the endowment (of 200 cents) bet in the lottery for the two treatments. To ease comparison, we take the average percentage of endowment bet in blocks of three rounds. These averages and the corresponding standard deviations (across individuals) are presented in Table 1. The final row of Table 1 gives the average percentage of endowment bet over all rounds.

The results display a clear treatment effect. In each round average bets are larger for treatment L than for treatment H. To determine the significance of the differences, we use the nonparametric Mann-Whitney test.[10] The final column reports *z*-values, which are a transformation of the Mann-Whitney *U*-value corrected for the presence of ties. These *z*-values

[9]In fact, after part 2 was finished, there was a short supplementary part in the experiment. In this part we tried to obtain additional information about subjects' risk preferences. This part, however, is not directly relevant to the present test.

[10] We cannot use the parametric t-test. This test assumes that the observations come from a normal Distribution, which is not possible, given the lower- and upper- bound of 0 and 100, respectively. Also, a Kolmogorov-Smirnov test rejects the hypothesis that the observations are from a normal distribution.

would suggest that subjects are, at least to a substantial extent, forward looking when they evaluate ("mentally account") risky decisions.

This seems to be in line with Benartzi and Thaler, who formulate MLA in terms of "prospective" utility. An additional hypothesis could be that experiencing (not just anticipating) gains and losses affects subjects' risk behavior. For such a backward looking hypothesis we find no support. In the course of the experiment, the subjects in Treatment H experience losses more frequently than do the subjects in Treatment L. If this were a driving force behind the difference between the two treatments, then we would expect this difference to be stronger in the final round(s) than in the first round(s). No support for this is found in the data. Moreover, we find no effect of different experiences with gains and losses between subjects.[12] The fraction of endowment bet is not significantly affected by subjects' experiences with the occurrence of gains and losses in the preceding round(s).

In part 2, subjects' endowments were again identical across rounds, but contrary to part 1, they differed across individuals. In each of the three rounds, a subject's endowment was equal to 1/3 of his or her earnings (W) from part 1 of the experiment ($S = W/3$). As a consequence, for each subject we have two variables of interest: first, the absolute amount bet, $Y := \sum_{t=10,11,12} X_t$ ($\leq W$), where for Treatment L we have $X_{10} = X_{11} = X_{12}$, and, second, the percentage of the endowment bet in the lottery, $F := 100Y/W$. The averages of both variables are presented in the first two rows of Table 2.

[12] For example, within each treatment we compared the bets of the subjects who had just experienced a gain with those who had just experienced a loss. If subjects were backward looking, we would expect the bets to be higher for the former than for the latter group. The effect is in the other direction, however. Although this finding is statistically insignificant, bets are larger after a loss than after a gain.

It appears that the treatment effect is in the same direction as in part 1. On average, subjects in Treatment L bet more in the risky lottery than do their counterparts in Treatment H. Both in absolute and relative terms, bets were larger if subjects were supplied with less information feedback and less freedom of choice. For the amount bet (*Y*), the difference is again highly significant. For the percentage of endowment bet (*Y*), the difference between the two treatments is less pronounced but still (marginally) significant. As the final row of Table II indicates, the increased willingness to take risks also pays off. Average total earnings (parts 1 and 2) in Treatment L are significantly larger than those in Treatment H.

	Treatment H[a]	Treatment L[a]	Mann-Whitney z[b]
Amount bet (Y)	707.3 (614.5)	887.1 (662.1)	-2.14 [0.016]
Percentage bet (F)	39.0 (30.0)	48.9 (32.1)	-1.62 [0.053]
Total earnings (parts 1 and 2)	1822 (1015)	2134 (745)	-1.78 [0.038]

Table 2: Average amount bet, average percentage bet, and average total earnings. a. # obs. = 41 (42) for treatment H (L). Standard deviations are in parentheses and, b. One-tailed significance levels (*p*-values) are in brackets.

4. Conclusions

This paper presents a direct experimental test of the prediction of myopic loss aversion (MLA), that a longer evaluation period makes a risky option with positive expected return look more attractive. Our results strongly support this prediction. We manipulated the evaluation period of one group of experimental subjects by giving them less information feedback and less freedom of adjustment than a control group. This manipulation was intended to make subjects evaluate risky financial investments in a more aggregated way. As a consequence, they are less likely to be deterred by the occurrence of losses. In particular, we observe higher earnings for the subjects who evaluate their investment in a more aggregate way. The results provide support for Benartzi and Thaler's (1995) explanation of the equity premium puzzle.

The results may also have practical relevance. Manipulating the evaluation period of prospective clients could be a useful marketing strategy for fund managers. Our results suggest that providing investors with less frequent information feedback about how a particular risky fund is doing might make the fund appear more attractive by decreasing the likelihood that a loss will be experienced. Similarly, if investors are given less freedom of adjustment ("tying their hands"), this may induce them to evaluate financial outcomes in a more aggregated way, and help them to resist the temptation to drop out after the occasional setback.

Of course, our experiment is highly stylized. For example, the subjects in the experiment only face risk (known probabilities of possible outcomes), whereas real-life investors mainly deal with uncertainty (unknown probabilities). Another issue is that our experiment took less than an hour, whereas the time elapsing between real investment decisions is usually much longer. Furthermore, the financial stakes for the experimental subjects are low compared with those of most real-world decision-makers. These features are a cause for caution in extrapolating of the results. They also suggest lines along which to pursue further experimental work.

Appendix: excerpt of instructions

(Translated from Dutch; full instructions are available upon request.)

Introduction [Read aloud only]

Welcome to our experimental study of decision-making. The experiment will last about 40 minutes. The instructions for the experiment are simple, and if you follow them carefully, you can earn a considerable amount of money. All the money you earn is yours to keep, and will be paid to you, privately and in cash, immediately after the experiment.

The experiment will consist of three parts. The instructions for the second part will be distributed to you after the first part has been finished. The instructions for part 3 will be announced at the completion of part 2. Before we start the experiment, however, you will be asked to pick one envelope from this pile. In the envelope you will find your so-called Registration Form. This form will be used to register your decisions and earnings. One of you, however, will find the announcement 'assistant' in the envelope. This person will assist us during the experiment, and will receive a payment that is equal to the average earnings of the other participants in the experiment.

On top of your Registration Form you will find your registration number This number indicates behind which table you are to take a seat. A separate table is reserved for the assistant. When everyone is seated, we will go through the instructions of part 1 of the experiment. After that, you will have the opportunity to study the instructions on your own, and to ask questions. If you have a question, please raise your hand, and I will come to your table. Please do not talk or communicate with the other participants during the experiment.

Are there any questions about what has been said until now? If not, then will the person on my left please be the first to pick an envelope, open it, and take the corresponding seat.

Instructions for part 1 [Treatment H; Read aloud and distributed]

Part 1 of the experiment consists of 9 successive rounds. In each round you will start with an amount of 200 cents (*f*2). You must decide which part of this amount (between 0 cents and 200 cents) you wish to bet in the following lottery.

> You have a chance of 2/3 (67%) to lose the amount you bet and a chance of 1/3 (33%) to win two and a half times the amount you bet.

You are requested to record your choice on your Registration Form. Suppose that you decide to bet an amount of X cents ($0 \leq X \leq 200$) in the lottery. Then you must fill in the amount X in the column headed *Amount in lottery,* in the row with the number of the present round.

Whether you win or lose in the lottery partly depends on your personal *win letter*. This letter is indicated on the top of your Registration Form. Your win letter can be A, B, or C, and is the same for all 9 rounds. In any round you win in the lottery if your win letter matches the *round letter* that will be drawn by the assistant, and you lose if your win letter does not match the round letter.

The round letter is determined as follows. After you have recorded your bet in the lottery for the round, the assistant will, in a random manner, pick one letter from a box containing three letters: A, B, and C. The letter drawn is the round letter for that round. If the round letter matches your win letter you win in the lottery; otherwise you lose. Since there are three letters, one of which matches your win letter, the chance of winning in the lottery is 1/3 (33%) and the chance of losing is 2/3 (67%).

Hence, your earnings in the lottery are determined as follows. If you have decided to put an amount of X cents in the lottery, then your earnings in the lottery for the round are equal to $-X$ if the round letter does not match your win letter (you lose the amount bet) and equal to $+2.5X$ if the round letter matches your win letter (you win two and a half times the amount bet).

The round letter will be shown to you by the assistant. You need to record this letter in the column *Round letters,* under win or lose, depending on whether the round letter does or does not match your win letter. Also you need to record your earnings in the lottery in the column *Earnings in lottery.* Your total earnings for the round are equal to 200 cents (your starting amount) plus your earnings in the lottery. These earnings are recorded in the column *Total earnings,* in the row of the corresponding round. Each time we will come by to check your Registration Form.

After that, you are requested to record your choice for the next round. Again you start with an amount of 200 cents, a part of which you can bet in the lottery. The same procedure as described above determines your earnings for this round. It is noted that your private win letter remains the same, but that for each round a new round letter is drawn by the assistant. All subsequent rounds will also proceed in the same manner. After the last round has been completed, your earnings in all rounds will be totaled. This amount determines your total earnings for part 1 of the experiment. Then the instructions for part 2 of the experiment will be announced.

Instructions for part 1 [Treatment L; Read aloud and distributed]

Part 1 of the experiment consists of 9 successive rounds. In each round you will start with an amount of 200 cents (ƒ2). You must decide which part of this amount (between 0 cents and 200 cents) you wish to bet in the following lottery.

> You have a chance of 2/3 (67%) to lose the amount you bet and a chance of 1/3 (33%) to win two and a half times the amount you bet.

You are requested to record your choice on your Registration Form. Suppose that you decide to bet an amount of X cents ($0 \leq X \leq 200$) in the lottery. Then you must fill in the amount X in the column headed *Amount in lottery.* Please note that you fix your choice for the next three rounds. Thus, if you decide to bet an amount X in the lottery for round 1, then you also bet an amount X in the lottery for rounds 2 and 3. Therefore, three consecutive rounds are joined together on the Registration Form.

Whether you win or lose in the lottery partly depends on your personal *win letter*. This letter is indicated on the top of your Registration Form. Your win letter can be A, B, or C, and is the same for all 9 rounds. In any round you win in the lottery if your win letter matches the round *letter* that will be drawn by the assistant, and you lose if your win letter does not match the round letter.

The round letter is determined as follows. After you have recorded your bet in the lottery for the next three rounds, the assistant will, in a random manner, for each of the next three rounds pick one letter from a box containing three letters: A, B, and C. For each of the three rounds a letter is drawn from a different box. The three letters drawn are the round letters for the present three rounds. If the round letter matches your win letter, you win in the lottery; otherwise you lose. Since each box contains three letters, one of which matches your win letter, the chance of winning in the lottery in a given round is 1/3 (33%) and the chance of losing is 2/3 (67%).

Hence, your earnings in the lottery for the three rounds are determined as follows. If you have decided to put an amount of X cents in the lottery, then your earnings in the lottery are equal to $-X$ for each round letter that does not match your win letter (you lose the amount bet for the round) and equal to $+2.5X$ for each round letter that matches your win letter (you win two and a half times the amount bet for the round).

The three round letters will be shown to you by the assistant. You need to record these letters in the column *Round letters,* under win *or* lose, depending on whether the round letter does or does not match your win letter. You also need to record your earnings in the lottery in the column *Earnings in lottery.* Your total earnings for the three rounds are equal to 600 cents (three times your starting amount of 200 cents) plus your earnings in the lottery. These earnings are recorded in the column *Total earnings,* in the row of the corresponding rounds. Each time we will come by to check your Registration Form.

After that, you are requested to record your choice for the next three rounds (4-6). For each of the three rounds you again start with an amount of 200 cents, a part of which you can bet in the lottery. The same procedure as described above determines your earnings for these three rounds. It is noted that your private win letter remains the same, but that for each round a new round letter is drawn by the assistant. The subsequent three rounds (7-9) will also proceed in the same manner. After the last round has been completed, your earnings in all rounds will be totaled. This amount determines your total earnings for part 1 of the experiment. Then the instructions for part 2 of the experiment will be announced.

References

Benartzi, Shlomo, and Richard Thaler (1995) "'Myopic Loss Aversion and the Equity Premium Puzzle," *Quarterly Journal of Economics* CX, 73-92.

Gollier, Christian, John Lindsey, and Richard Zeckhauser (19960 "'Investment Flexibility and the Acceptance of Risk," working paper, Gremaq, University of Toulouse.

Kahneman, Daniel, and Dan Lovallo (1993) "Timid Choices and Bold Forecasts: A Cognitive Perspective on Risk Taking," *Management Science*, XXXIX, 17-3 1.

Kahneman, Daniel, and Amos Tversky (1979) "Prospect Theory: An Analysis of Decision under Risk," *Econometrica*, XLVII, 263-9 1.

Kahneman, Daniel, and Amos Tversky (1984) "'Choices Values and Frames," *American Psychologist,* XXXIX, 341-50.

Mehra, Rajnish, and Edward Prescott (1985) "The Equity Premium: A Puzzle," *Journalof Monetary Economics*, XV, 145-61.

Samuelson, Paill (1963) "Risk and Uncertainty: A Fallacy of Large Numbers," *Scientia*, XCVIII, 108-13.

Thaler, Richard (1985) "Mental Accounting and Consumer Choice," *Marketing Science,*IV, 199-214.

Thaler, Richard, Amos Tversky, Daniel Kahneman, and Alan Schwartz (1997) "The Effect of Myopia and Loss Aversion on Risk Taking: An Experimental Test," *Quarterly Journal of Economics,* CXII.?

Tversky, Amos, and May'a Bar-Hillel (1983) "Risk: The Long and the Short," *Journal of Experimental Psychology: Learning, Memory, and Cognition*, IX, 713-17.

Tversky, Amos, and Daniel Kahneman (1992) "Advances in Prospect Theory: Cumulative Representation of Uncertainty," *Journal of Risk and Uncertainty*, V, 297-323.

Chapter 2

Probability Judgments in Multi-Stage Problems: Experimental Evidence of Systematic Biases[1]

Abstract: We report empirical evidence that in problems of random walk with positive drift, bounded rationality leads individuals to underestimate the probability of success in the long run. In particular, individuals who were given the stage-by-stage probability disterbution failed to aggregate this information in a multi-stage case. Estimations of the long-run probability distribution did not differ much from the given stage-by-stage probability distribution, and were systematically lower than the accurate one. Applications to risk perception in financial markets are considered.

---

[1] This paper was first published in *Acta Psychologica* 93 (1996) 59-68.

1. Introduction

The standard approach to sequential decision making under uncertainty (i.e. the Savage (1954) subjective expected utility theory) assumes that people are indifferent to the way problems are set, and are only interested in the probability distribution over final outcomes (see Hammond, 1988; Machina, 1989). In particular, people are assumed to follow the reduction of compound lotteries axiom, stating that a multi-stage lottery is equally attractive as the one-stage lottery that yields the same prizes with the corresponding multiplied probabilities. For example, consider the following two lotteries: in the first, a fair coin is tossed twice in a row. If it falls on its head twice or on its tail twice, the decision maker wins $1; he loses $1 if it falls once on each side. In the second lottery, two fair coins are tossed at the same time, and the payoffs are the same as in the first lottery. The reduction axiom states that the decision maker is indifferent between the first and the second lottery.

While there exists a literature following Kahneman and Tversky (1979) (see Camerer, 1995, for a short survey) that questions this axiom, a common assumption is that people are capable of accurately estimating the reduced probabilities of compound lotteries, or at least that mistakes are not systematic, and estimations are accurate on average. This is surprising because it is not difficult to construct sequential problems in which, for bounded rationality reasons, people fail to estimate reduced probabilities (see the book edited by Kahneman et al., 1982, or the works of Bar-Hillel, 1973, and Wagenaar and Sagaria, 1975, which are described below).

This paper looks at a different aspect of this problem, namely processes of random walk with positive drift, which are very important in many real-life economic decisions. We test whether probability judgment is 'good' in this kind of environment, or is it systematically biased - and if so how. Versions of the following investment game[2] are used:

> An option on the price of a stock is for sale. Today the price of the stock is \$$x$, and every day it either goes up or down by \$ 1, with probability p and $1 - p$ respectively. The option will be realized and pay \$0 if the price of the stock will reach \$0, and \$$n$ if the price will reach \$$n$. What is the probability that the realization price will be \$$n$?

Results of three experiments with the game are reported, in which we controlled for the following three parameters:

1. we changed the starting amount to x = \$3, \$5, and \$7, fixing $p = 0.6$ and $n = 10$,
2. we changed the size of the interval to $n = 4$, 6, 10, and 14, when $x = n / 2$ and $p = 0.6$, and
3. for $x = 5$ and $n = 10$, we changed the stage-by-stage probability of success to $p = 0.55$, 0.58, 0.6, 0.65, and 0.7.

The results suggest that people use the stage-by-stage probability as an anchor, and adjust insufficiently. Estimations are biased toward the direction of the stage-by-stage probability, resisting in underestimation of the overall probability of success. One consequence is that while individuals do quite well in estimating the probabilities in 'small' intervals, in which the compound probability does not differ much from the stage-by-stage probability, they fail to appreciate the affect of enlarging the interval i.e., the fact that the probability of success

[2] A similar game, known as "The gambler's ruin problem", is a classical problem in the random walk literature. Early solutions by Bernoulli and De Moivrc are described in Thatcher (1957). For detailed solutions see Ross (1989). This literature was not concerned with the bounded rationality aspects of this problem.

increases. For that reason, for the values of n tested, underestimations increased with n. Another consequence is that subjects fail to fully appreciate increases in the stage-by-stage probability, i.e. the fact that a 'small' increase in the stage-by-stage probability implies a 'large' increase in the overall probability of success.

In the paper we try to get some insight into the relevance of this to 'real-life' problems, such as the equity premium puzzle. It may be that the failure of traditional risk measures to explain behavior in many cases is not a case of a bad theory of risk behavior, as much as a simple misjudgment of the objective probabilities by people. See Arrow (1982).

2. Computing the compound probability of success

Let $p_x(t)$ be the probability of getting n after t stages for a player who starts with x. Denote the infinite case by p_x , i.e. $p_x = \lim_{t \to \infty} p_x(t)$. The probabilities p_x satisfy the following system of equations:

$$\begin{aligned} p_n &= 1 \\ &\vdots \\ p_x &= (1 - p) \times p_{x-1} + p \times p_{x+1} \qquad \text{for } 0 < x < n, \\ &\vdots \\ p_0 &= 0 \end{aligned}$$

Proposition 2.1. The explicit solution of the system, for $p \neq 0.5$ is:

$$P(x) = \frac{[(1 - p)\, p^{-1}]^{x} - 1}{[(1 - p)\, p^{-1}]^{n} - 1}$$

Proof. This is a system of $n + 1$ linear equations in $n + 1$ unknowns (p_0, ..., p_n). It is easily seen that the detenninant of the system is non-zero, hence, the system has at most one solution. Direct verification shows that the equation in the proposition is a solution to the system. Hence it must be the unique solution. □

In Table 1 are the p_x's for a few different x's and p's. $n = 10$.

	$p = 0.5$	$p = 0.55$	$p = 0.58$	$p = 0.6$	$p = 0.65$	$p = 0.7$
P_{10}	1	1	1	1	1	1
P_9	0.9	0.96	0.98	0.99	0.998	0.9997
P_8	0.8	0.91	0.96	0.98	0.995	0.9990
P_7	0.7	0.86	0.93	0.96	0.989	0.9976
P_6	0.6	0.79	0.89	0.93	0.978	0.9940
P_5	0.5	0.71	0.83	0.88	0.957	0.9857
P_4	0.4	0.62	0.75	0.82	0.918	0.9670
P_3	0.3	0.51	0.65	0.72	0.850	0.9200
P_2	0.2	0.37	0.5	0.57	0.710	0.8200
P_1	0.1	0.20	0.29	0.34	0.460	0.5700
P_0	0	0	0	0	0	0

Table 1: Values of p_x are the probabilities of reaching $n = 10$, starting with x, when p is the stage-by-stage probability

3. Method and results

In this section, experimental results from three experiments are presented, showing underestimation.

3.1 Experiment 1: Changing the starting amount with experienced subjects

We wanted to cheek whether having some 'experience' with the game will make the estimates more accurate. First year students in economics at Tilburg University participated in the experiment. The students played a version of the game (see Appendix A) for real money.[3]

[3] We used Dutch guilders, scaling the game such that the change in prize in every stage was f 2.5 instead of $1, e.g. when we say that the game started with $5 it actually started with f 12.5 (at the time f 2.5 = $1.6). To reduce confusion, we continue presenting the results in dollars.

Each student played privately and independently of the others. In total 28 subjects participated; 10 students started with \$3, 10 with \$5 and 8 students with \$7. The probability used was p = 0.58, and n = 10. Each session took at most 30 minutes. After playing, subjects were asked indirectly (see Appendix B) about their estimations. This was done in order to check whether the mistakes resulted from confusion created by terms like 'Probability' and 'chance'. Their responses are presented in Table 2.

Subject	Start with \$3	Subject	Start with \$5	Subject	Start with \$7
1	0.85	11	0.95	21	0.85
2	0.75	12	0.80	22	0.83
3	0.70	13	0.70	23	0.80
4	0.62	14	0.70	24	0.75
5	0.58	15	0.58	25	0.58
6	0.58	16	0.58	26	0.58
7	0.58	17	0.58	27	0.58
8	0.40	18	0.58	28	0.44
9	0.25	19	0.37		
10	0.05	20	0.30		
Mean	0.536	Mean	0.614	Mean	0.676
Actual p_x	0.646	Actual p_x	0.834	Actual p_x	0.933

Table 2: The right column in each starting amount gives the estimated p_x for subjects who first played the game fond real money. p = 0.58 and n = 10. For each x, subjects are ordered by their estimation

Twenty-four out of the 28 subjects underestimated p_x. This first experiment shows that underestimation exists, even after some experience in playing. An interesting observation is that, on average, playing more stages (more 'experience') did not result in more accurate estimations. Another experiment, not incentive motivated, used 16 seminar participants (professors and Ph.D students in economics) from Tilburg University as subjects. Each subject was asked to estimate p_x for 4 different x's. Although these subjects are not 'normal people', in the sense that they know more about probability theory than most people, 54 out of the 64 responses underestimated p_x.

3.2 Experiment 2: Changing the size of the interval

Most of the literature on sequential decision problems uses two stage lotteries as a sole representation of dynamics. This is done under the assumption that moving from one-stage to two-stage lotteries captures the essential aspects of dynan-iics, and moving from two-stage to multi-stage lotteries is trivial. We show that in our random walk example, there is no 'irrationality' in a two-stage set-up, but subjects become 'more irrational' with every stage added.

To do this, different sizes of intervals were used, keeping the rest of the rules the same. In Table 3 are the reduced probabilities for different values of n. when starting with $x = n/2$ (calculated using Proposition 2. 1).

n	2	4	6	8	10	12	14	16	18	20	50
$P_n/2$	0.60	0.69	0.77	0.84	0,88	0.92	0.94	0.96	0.97	0.98	0.99996

Table 3: The probabilities of reaching different values of n, when $p = 0.6$, and starting at $n/2$

As one can see, the reduced probabilities converge to 1 very rapidly. The question raised now is, would subjects, although underestimating the reduced probabilities, understand that they converge to 1?

We used undergraduate students in economics, and gave them monetary incentives to find the accurate probabilities. We had 4 groups of subjects, one with 4 subjects, one with 5, and 2 with 6 subjects each. Appendix C is an example of a questionnaire for $n = 4$. The results are presented in Table 4.

Subject	$n = 4$	Subject	$n = 6$	Subject	$n = 10$	Subject	$n = 14$
1	0.8	1	0.8	1	1	1	0.94
2	0.7	2	0.67	2	0.8	2	0.8
3	0.6	3	0.6	3	0.65	3	0.6
4	0.5	4	0.6	4	0.6	4	0.6
		5	0.6,	5	0.35	5	0.6
		6	0.36			6	0.6
Mean	0.65	Mean	0.605	Mean	0.68	Mean	0.69
Actual p_x	0.69	Actual p_x	0.77	Actual p_x	0.88	Actual p_x	0.94

Table 4: Estimations of $P_{n/2}$ for different values of n

The underestimates are robust even under this treatment (16 out of 21 subjects, but 9 out of 11 for n = 10 and n = 14), and the size and frequency of underestimations increase with n. Another observation is that the mean of the observations, within the range of $4 \leq n \leq 14$, did not converge to 1, i.e. the expectations are not monotonic and do not differ much from each other.

Subject	Prob 0.55	Subject	Prob 0.6	Subject	Prob 0.65	Subject	Prob 0.7
1	0.90	16	0.99	31	0.90	46	0.99
2	0.68	17	0.80	32	0.85	47	0.83
3	0.59	18	0.77	33	0.82	48	0.82
4	0.59	19	0.73	34	0.80	49	0.70
5	0.57	20	0.70	35	0.74	50	0.70
6	0.55	21	0.69	36	0.70	51	0.70
7	0.55	22	0.65	37	0.65	52	0.70
8	0.55	23	0.64	38	0.65	53	0.70
9	0.55	24	0.60	39	0.65	54	0.70
10	0.55	25	0.60	40	0.65	55	0.65
11	0.55	26	0.60	41	0.60	56	0.60
12	0.55	27	0.52	42	0.58	57	0.58
13	0.55	28	0.50	43	0.50	58	0.30
14	0.45	29	0.20	44	0.50	59	0.17
15	*	30	*	45	0.35	60	*
Mean	0.58	Mean	0.64	Mean	0.66	Mean	0.65
Actual	0.73	Actual	0.88	Actual	0.96	Actual	0.99

Table 5: Estimations of p_x for different values of p, x = 5 and n = 10. * Observations that did not add up to one.

3.3 Experiment 3: Changing the stage-by-stage probability

In this experiment we fixed $x = 5$ and $n = 10$, and varied p. We used 60 first year economics students, in four groups of 15 subjects. Each group was asked about one p. Instructions were similar to those of Experiment 2, but a different reward scheme was used (see Appendix D). The results are presented in Table 5.

Again we see underestimations of p_x (54 out of 57 subjects), and we see that changing p did not change the mean of the estimations which, apart from the case of $p = 0.55$, are almost identical. This implies that subjects were not sensitive to changes in p.

4. Discussion: Anchoring and adjustment heuristic

The evidence indicates that when estimating the compound probability of success (p_x), subjects use the stage-by-stage probability of success (p) as an anchor. Apparently, subjects 'start' with p, anchor to that, and either do not adjust at all, or adjust insufficiently to changes in the parameters. In total, 40 out of the 106 estimations were $p = p_x$. Moreover, if we look at a comparison of the distance $|(p_x - \text{mean})/(p - \text{mean})|$, as done in Table 6, we see that the mean of estimates was at least 2.5 times closer to p than to p_x for all but the $n = 4$ case in Experiment 2.

The consequence of changing the starting amount was tested in Experiment 1, where it was shown that there is some adjustment, always in the correct direction, but the adjustment is insufficient. In Experiment 2 we changed the size of the interval, and found no sign of adjustment. Experiment 3 shows that estimations are not sensitive to changes in the stage-by-stage probability.

	p	p_x	mean	$\|(p_{x'} - \text{mean})/(p - \text{mean})\|$
Experiment 1				
1	0.58	0.646	0.536	2.5
2	0.58	0.834	0.614	6.5
3	0.58	0.933	0.676	2.7
Experiment 2				
1	0.60	0.69	0.650	0.08
2	0.60	0.77	0.605	33
3	0.60	0.88	0.680	2.5
4	0.60	0.94	0.690	2.8
Experiment 3				
1	0.55	0.73	0.58	5
2	0.60	0.88	0.64	6
3	0.65	0.96	0.66	30
4	0.70	0.99	0.65	6.8

Table 6: Comparison of the distance $|(p_{x'} - \text{mean})/(p - \text{mean})|$ using the mean of estimations, p, and p_x from all the experiments

This is not the first attempt to look at estimations of compound probabilities. Bar-Hillel (1973) investigated the hypothesis that the subjective probability of compound events are systematically biased in the direction of their components, resulting in overestimation of the likelihoods of conjunctive events and underestimation of the likelihood of disjunctive events, e.g. a probability of a conjunctive event may be the probability of winning 5 times in a row, and the probability of a disjunctive event is the stage-by-stage probability. Bar-Hillel concluded that "... The probability of the individual stage in a chain of events thus appears to have greater influence on the evaluation of the whole chain's probability than the number of stages in question" (p. 405). This is similar to our result in the sense that people anchor to the probability of the individual stage, and fail to fully appreciate the affect of enlarging the number of stages. However, this work focuses on the probability of a certain path, and not on the probability of outcomes. Note also that increasing the number of stages has an opposite effect as compared to our story, i.e. it reduces the compound probability.

Other studies that report underestimation in multi-stage problems, are Wagenaar and Sagaria (1975), and Wagenaar and Tinuners (1979). These studies consider a different type of problem, namely estimations of exponential growth. They show that exponential growth is considerably underestimated; people tend to extrapolate exponentially, that is with a constant multiplier for successive steps, but with an exponent that is too small.

5. Application to financial markets

We showed that in a bounded random walk set-up, with positive drift, most subjects underestimate the reduced probabilities of reaching the upper bound. Why should this result interest economists? For example, the traditional finance literature assumes that asset prices in an efficient capital market follow a random walk with positive drift, i.e. that capital markets "have no memory" (Brealey and Myers, 1988, p. 289).[4] Our findings suggest that investors will fail to appreciate the difference in the retums between the short and the long run. For example, a stock that is traded daily, and whose price follows a random walk with a known 'very small' daily positive expected return, rnay do 'very well' in the long run, much better than people expect it to do on the basis of its daily performance. This implies that investors' perceived risk of that kind of asset is systematically higher than the objective risk and, as a result, assets are undervalued.

4 This approach is controversial nowadays (e.g. Fama, 1991, or De Bondt and Tbater, 1994). Yet as a first approximation it is still accepted, and that is enough for our case.

Another difficulty investors inay have is underestimating the chance that a slightly better asset (higher p) will accumulate much larger wealth in the long run.

For example the equity premium puzzle, which is the empirical fact that stocks have outperformed bonds over the last century in a way that is hard to explain with plausible levels of investor risk aversion (Mehra and Prescott, 1985), may partly be the result of investors' misjudgment of risk.

One obvious question is whether markets would 'fix' these underestimations. The problem is that, since the 'objective probabilities' in the stock market are unknown, there cannot be any empirical proof of the kind given in this paper. It may be that the price in a market will reflect the accurate probabilities even if (most) participants are not able to correctly estimate these probabilities. On the other hand, there is evidence that markets are not always efficient (e.g. De Bondt and Thaler, 1994). For an elaborated discussion about the role of risk perception in financial markets see Arrow (1982).

In future research, we would like to address this question, with the aim of tackling Camerer's challenge "Whether judgment and choice violations matter in markets is a question that begs for empirical analysis" (Camerer, 1987, p. 981).

Appendix A

A.1 Instructions for subjects who played the game for real money

In a few minutes we will give you $5 and ask if you want to participate in a game in which at every step you can either win or lose $1. The chance of winning $1 is 58% and the chance of losing $1 is 42%. If you choose to leave the game, you can stop and take the $5. If you decide to play, you will either have $4 or $6 after the first stage. Then, you can leave the game with your money, or participate in the next stage, in which, again, you can either win or lose $1 with the same chances as before. The game goes on until either you choose to stop, or your money reaches {$0 or $10}, or after 100 stages.[5]

Appendix B

B.1 Indirect method for finding estimations (note that the question is phrased such thal it is equivalent to the initial problem)

Say that we take 100 students and let them play the game, with one difference: they will have to play till $0 or $10. Can you guess how many of theni will end up with $0 and how many with $10?

[5] The probability that the gaine will not end within 100 stages is less than 0.001, hence p_x is relevant even for $t = 100$. For a discussion of the use of this restriction see Gneezy (1995).

Appendix C

C.1 Estimations for different sizes of intervals with monetary incentives (for n = 4)

Please answer the following question, which is also given to other students in the room. After all of you have finished answering, we will collect and check the answers. We will find the best answer and give $10 to the student who gave it. If snore than one student gives the best answer, we will split the money between the students who gave this answer.

The game: Mr. X is given $2 and then a series of lotteries take place. In each lottery he either wins or loses $1. The chance of winning $1 is 60% and the chance of losing $1 is 40%. So, after the first lottery, Mr. X will either have $3 (with 60% chance) or $1 (with 40% chance), and so on. The lotteries will be conducted till Mr. X will either have $0 or $4.

The problem: What do you think is the chance that Mr. X will:

(a) finish with $4?

(b) finish with $0?

Appendix D

D.I you will be paid according to the following rule

you will start with $15, and for every 1% of 'mistake' $1 will be deduced from your payoff. The mistake is the absolute value of [your guess (in percentages) minus the actual chance]. For example, if you guess accurately, you get $15. If you make a 10% mistake (either overestimate or underestimate), you get $5. If your mistake is bigger than or equal to 15% you will not be paid at all.

References

Arrow, K.J. (1982) "Risk Perception in Psychology and Economics," *Economic Inquiry*, 20, 1-9.

Bar-Hillel, M. (1973) "On the Subjective Probability of Compound Events," *Organizational Behavior and Human Performance* 9, 396-406.

Brealey, R. and S. Myers (1989) *Principles of Corporate Finance* (3rd ed.). New York: McGraw-Hill.

Camerer, C. (1987) "Do Biases in Probability Judgment Matter in Markets? Experimental Evidence," *American Economic Review*, December, 991-997.

Camerer, C., "Individual Decision Making", In: J.H. Kagel and A.E. Roth (Eds.), *Handbook of Experimental Economics*. Princeton, NJ: Princeton University Press, 1995.

De Bondt, W. and R.H. Thaler (1994) "Financial Decision-Making in Markets and Firms: A Behavioral Perspective," In: R. Jarrow, V. Maksimovic, and W.T. Ziemba (Eds.), Amsterdam: North-Holland.

Fama, E.F. (1991) "Efficient Capital Markets: II", *Journal of Finance* 46, 1575-1617.

Gneezy, U. (1995) "Updating the Reference Level: Experimental Evidence", Draft.

Hammond, P.J. (1988) "Consequentialist Foundations for Expected Utility. Theory and Decision 25, 25-78.

Kahneman, D., P. Slovic and A. Tversky (1982) *Judgment Under Uncertainty: Heuristics and Biases*. New York: Cambridge University Press.

Kahneman, D. and A. Tversky (1979) "Prospect Theory: An Analysis of Decision Under Risk," *Econometrica* 47, 263-291.

Machina, M.J. (1989) "Dynamic Consistency and Non-Expected Utility Models of Choice Under Uncertainty," *Journal of Economic Literature* 27, 1622-1668.

Mehra, R. and E.C. Prescott (1985), "The Equity Premium; A Puzzle," *Journal of Monetary Economics* 15, 145-161.

Ross, S.M. (1989) *Introduction to Probability Models*. London: Academic Press.

Savage, L.J. (1954) *The Foundation of Statistics*. New York: Wiley.

Thatcher. A.R. (1957) "Studies in the History of Probability and Statistics, VI. A Note on the Early Solutions of the Duration of the Play," *Biometrika* 44, 515-518.

Wagenaar, W.A. and S.D. Sagaria (1975) "Misperception of Experimental Growth," *Perception and Psychophysics* 18, 422-426.

Wagenaar, W.A. and H. Timmers (1979) "The Pond-and-Duckweed Problem," *Acta Psychologica* 43, 239-251.

Chapter 3

Experimental Investigation of Perceived Risk in Finite Random Walk Processes[1]

Abstract: The hypothesis that, on average, people accurately estimate probabilities in random walk processes is experimentally investigated. Individuals are confronted with a process that starts with $X, and in every stage either goes up or down by $1, with probabilities p and 1 - *p* respectively. For different values of *p*, individuals were asked to estimate what is the chance that after 10 stages the system will be at a point higher than or equal to $*X*. Systematic mistakes in estimations were observed. In particular, estimations were centered around the stage-by-stage probability (*p*) rather then around the actual probability. We connect this finding to recently developed explanations of the equity premium puzzle.

[1] This is a joint project with Marcel Das

1. Introduction

There is a large psychological literature concerning the way individuals estimate probabilities. In particular, many systematic biases are documented (see e.g. the book edited by Kahneman et al. [1982]). This line of research was criticized recently. See the discussion in Kahneman and Tversky [1996] and Gigerenzer [1996] and the references there. One of the reasons of the critique is that the relevance of these studies to economic problems is not always clear. In this paper we consider probability estimation in random walk processes. These type of processes are of great importance in finance, since they are assumed to describe price changes in so-called efficient markets (see Section 4). Clearly, systematic mistakes in probability estimation will lead to systematic mistakes in risk perception. This may be helpful in understanding observations from real markets, and in developing a better behavioral theory.

We experimentally investigate whether people assess the probability of several outcomes after N periods correctly, and if not, are the mistakes systematically 'optimistic' (pessimistic), i.e. viewing the process as less (more) risky than it really is. At this point no explicit definition of risk is given; it will be clear in the context of the experiment. We use the following setup:

> An investor is given a stock that is worth \$$X$. Then a process of N stages begins. In every stage, the price of the stock either goes up or down by \$1, with probability p and $1 - p$ respectively. If the price of the stock reaches \$0 within N stages, then the stock will be worth \$0 for ever.
>
> Problem: What do you think is the probability that the stock will be worth at least \$$X$ after N stages (p_x)?

We investigated this game with p = 0.25, 0.33,0.4, 0.6, 0.67, and 0.75. In all treatments $N = 10$ and $X = 3$. 104 subjects participated, each presented with only one p. The subjects were awarded according to the accuracy of their answers (see the Appendix).

We find support for the hypothesis that estimations of the compound probabilities are centered around the stage-by-stage probability. We call this "anchoring heuristic". This heuristic explains the under-estimations of the compound probabilities for $p > 0.5$ and over-estimations for $p < 0.5$.

Benartzi and Thaler [1996] investigated a similar problem from a different point of view. In their study subjects make several choices between a certain amount and a gamble. The choices vary in the way the gambles are described. In some cases the gamble is described as N plays of gamble X; in other cases as an explicit distribution of possible outcomes. Benartzi and Thaler find that subjects turn down the gamble when presented in the repeated trial format, and accept it in the distribution format. They argue that a partial explanation of this is that subjects do not give nearly enough weight to the value of the repeated trials (much like in the "Law of Small Numbers" of Kahneman and Tversky [1996]; see the book edited by Kahneman et al., [1982]).

In the literature on probability assessment there is extensive discussion on problems connected with applications of Bayesian methods. Relatively little has been done about the practical problem of assessing prior distributions. For early references, see Stael von Holstein [1970, Chapter 10], Bar-Hillel [1973] and Wagenaar and Sagaria [1975]. For a more recent study, see Gneezy [1996]. The purpose of the current study is twofold: first to increase our understanding of the psychological process that leads to the assessment of compound probabilities, and second, to use this knowledge to understand an economical phenomenon: the equity premium puzzle.

The paper proceeds as follows. Section 2 states the hypotheses to be tested, illustrates the computation of p_x, and ends with describing the method used. Section 3 gives the results, and Section 4 contains some discussion of the results.

2. Hypotheses, computation of p_x, and method

2.1 Hypotheses

A traditional assumption in finance is that even if, for bounded rationality reasons, not all individuals estimate probabilities accurately, there are no systematic mistakes in the estimations. So the benchmark hypothesis we use is:

H1: The median of the individuals' estimations is p_x.

Against this benchmark hypothesis, we test the hypotheses:

H2: For a given p, the majority of individuals over-estimate p_x.

H3: For a given p, the majority of individuals under-estimate p_x.

Finally, we test the anchoring hypothesis:

H4: The median of the answers is the stage-by-stage probability p.

2.2 Computation of p_x

To calculate p_x we define the *transition probability matrix* P with elements p_{ij}:

$$P = [p_{ij}], \quad i,j = 1, 2, \dots, X + N + 1,$$

where the elements p_{ij} are given by

$$p_{ij} = \begin{cases} 1 & \text{if } i = j = 1, \\ p & \text{if } i = 2,\dots, X + \mathrm{N}, \text{ and } j = i + 1, \\ 1 - p & \text{if } i = 2,\dots, X + N + 1, \text{ and } j = i - 1, \\ 0 & \text{otherwise.} \end{cases}$$

If we define $q = 1 - p$, then P can be written as

$$P = \begin{bmatrix} 1 & 0 & & \cdots & & 0 \\ q & 0 & p & & & \\ & q & 0 & p & & \vdots \\ \vdots & & \ddots & \ddots & \ddots & \\ & & & q & 0 & p \\ 0 & & \cdots & & q & 0 \end{bmatrix}$$

Note that it is enough to have X + N + 1 rows, since after N stages the stock can be worth at most \$ $(X + N)$. The probabilities of reaching a value of \$$j$ ($j = 0, \dots, X + N$) after N stages, when the starting amount is equal to \$$X$, can be found in the $(X + 1)$-th row of $Q = [q_{ij}] = P^N$. Finally, the p_x is then calculated by

2.3 Method

We fixed $X = 3$ and had 6 treatments, with the stage-by-stage probability (p) of 0.25, 0.33, 0.40, 0.60, 0.67, 0.75 in treatment 1, . . . , 6 respectively. Altogether 104 subjects participated: 18, 15, 18, 17, 18, 18 in treatment 1,. . . , 6 respectively. Subjects were first year students in economics at the second semester of their studies. Most of them had participated successfully in a basic statistics course. After a short introduction, subjects received the instruction (see the Appendix for the instructions for treatment 1). The answers of subjects were evaluated and rewarded as described in the instructions. The experiment took 20 minutes (excluding the paying time).

3. Results

The answers given by the subjects are presented in Table 1. We first test the benchmark hypothesis (H1). It is not wise to use a test based upon the mean of the observations, since this test will be very sensitive to outliers (in particular with this kind of number of observations, see e.g. Hampel et al. [1986]). A sign test is an alternative.[2] Under H1 the number of individuals that under-estimates should equal the number of individuals that over-estimates the compound probability p_x. From Table 2 we can see that hypothesis H1 is rejected for $p = 0.25$, 0.67, and 0.75, and is not rejected for $p = 0.33$, 0.40, and 0.60 (we reject H1 when one of the two probabilities is lower than 2.5%).[3] We should note that for $p = 0.25$ there are some (unreasonable) high answers. If we drop these observations from the analysis, the number of

[2] Under the null hypothesis, the number of respondents that under-estimates p_x follows a binomial distribution $B(n, q)$ with parameters n equal to the number of observations and q equal to .5.

[3] For $p = 0.25$ one respondent gave the exact answer. In that case the test is based upon the answers different from p_x (conditional sign test).

under-estimations is still larger than the number of over-estimation, but the result is not significant anymore.

p	.25	.33	.40	.60	.67	.75
p_x	.07	.19	.33	.78	.89	.96
# obs.	18	15	18	17	18	18
	.02	.04	.03	.35	.04	.04
	.05	.05	.04	.53	.33	.50
	.05	.08	.05	.55	.47	.50
	.06	.15	.05	.60	.50	.50
	.07	.15	.13	.60	.52	.62
	.10	.15	.20	.60	.60	.65
	.14	.20	.24	.66	.65	.65
	.15	.21	.25	.70	.65	.70
	.17	.22	.35	.75	.66	.73
	.19	.30	.38	.77	.67	.75
	.20	.49	.40	.80	.67	.80
	.25	.63	.48	.80	.68	.80
	.34	.65	.53	.80	.75	.83
	.85	.67	.65	.80	.80	.84
	.89	.90	.65	.80	.87	.90
	-90		.70	.85	.95	.90
	.93		.90	.85	.97	.90
	.99		.95		1.00	.99

Table 1 : Data for the six different stage-by-stage probabilities; p = stage-by-stage probability and p_x = compound probability.

When H1 is rejected, we then test whether H2 or H3 will be rejected. In the case of p = 0.25, we cannot reject hypothesis H2, but hypothesis H3 is rejected. For $p = 0.67$ and $p = 0.75$, H2 is rejected, but H3 is not rejected (see Table 2).

P	0.25	0.33	0.40	0.60	0.67	0.75
p_x	0.07	0.19	0.33	0.78	0.89	0.96
r	4	6	8	10	15	17
$P\{X \leq r\}$	0.0245	0.304	0.407	0.834	0.999	1.000
$P\{X \geq r\}$	0.994	0.849	0.760	0.315	3.77E-3	7.25E-5

Table 2 : Significance probabilities corresponding to Hl (and H2 and H3). X is the number of respondents that under-estimates p_x and r is realization.

Let's now consider the hypothesis H4. Again, we use a sign test: under H4, the number of individuals that give an answer below p equals the number of individuals that give an answer above p. From Table 3 we see that, for all treatments, we cannot reject hypothesis H4, i.e. estimations are centered around p (significance level is 2 x 2.5%).

P	0.25	0.33	0.40	0.60	0.67	0.75
p_x	0.07	0.19	0.33	0.78	0.89	0.96
r	11	10	10	3	9	9
$P\{X \leq r\}$	0.928	0.941	0.834	0.0287	0.773	0.685
$P\{X \geq r\}$	0.166	0.151	0.315	0.994	0.402	0.500

Table 3 : Significance probabilities corresponding to H4. X is the number of respondents that under-estimates p_x and r is realization.

In Figure 1 we graphically illustrate what we just found. The guesses of p_x are plotted for each treatment, together with the actual p_x as a function of p.

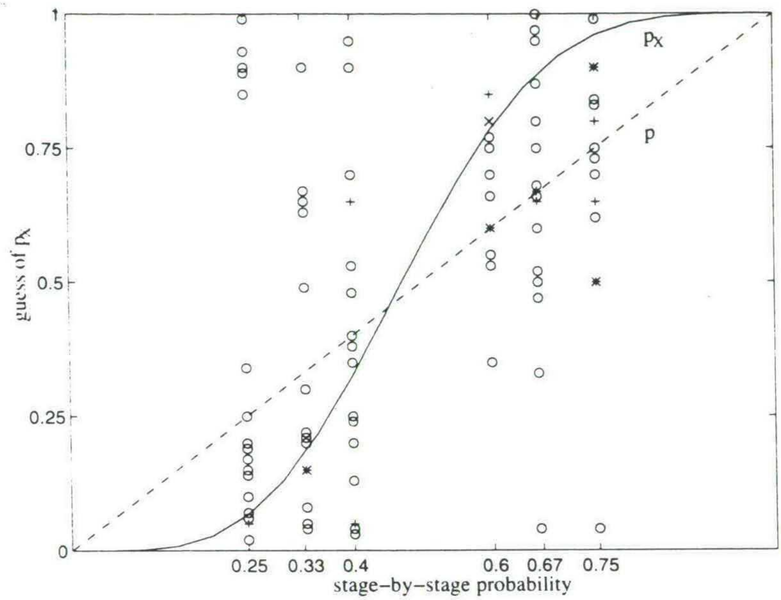


Figure 1 : Graphical representation of the guesses of p_x for each treatment together with the stage-by-stage probability (p, dashed line) and the compound probability (p_x, solid line). The '0' corresponds with one observation, '+' with two, '*' with three and 'x' corresponds with five observations.

As can be seen from this figure, for $p = 0.25$, 0.33, and 0.40, we have $p_x < p$, and for $p = 0.60$, 0.67, and 0.75 we have $p_x > p$. Since, as shown above, estimations are 'centered' around p and not around p_x, it is not surprising that in cases where $p_x < p$ the fraction of under-estimations is smaller than 0.5, and when $p_x > p$ it is larger than 0.5.

4. Discussion

Many tasks in life demand the use of heuristics, rather than explicit calculations. Some of these heuristics are important and very useful. However, heuristics may lead to biases in perception. Psychologists focus on the study of these biases in order to get more insight into the way heuristics are formed (see the discussion in Kahneman and Tversky [1996] and Glgerenzer

[1996]). In this paper we consider the assessment of probabilities in random walk processes. We finger out a heuristic that is used by people in assessing compound probabilities in these kind of set-ups, namely anchoring to the given stage-by-stage probability.[4]

We try to learn from the bias observed about the behavior in real financial markets. One of the most important hypotheses that has evolved from the research of financial markets, and has been empirically investigated, is the efficient markets hypothesis. According to this hypothesis, financial markets are 'efficient', and prices should reflect a rational forecast of the present value of future dividend payment. The efficient markets hypothesis has also been traditionally associated with the assertion that future price changes are unpredictable, although stock prices have a positive drift, see De Bondt and Thaler [1989, p.189]. The arbitrage forces are supposed to guarantee that prices adjust, and then move again, randomly, in response to unpredictable events. In a classical example, Fama [1965, p.98] writes: "It seems safe to say that this paper has presented strong and voluminous evidence in favor of the random walk hypothesis." Of course, this does not imply that all stocks follow the same random walk. The future prospects of stocks may still differ. These future prospects may create different random walk processes for different stocks. Looking N periods into the future, different processes typically imply different probability distributions over prices. The state of the art today is the belief that only to a first approximation, financial markets follow a random walk. For an elaborate discussion of the efficient markets and random walk hypothesis, see De Bondt and Thaler [1989] and Fama [1991].

[4] An important observation is that although people use 'wrong' heuristics, it seems to us that in the process of teaching students, virtually all the attention is given to developing their skills in solving problems analytically. We believe that not enough attention is given to the development of 'heuristic skill', e.g. the marginal contribution for an MBA student of a course which tries to develop heuristic skills may be very high.

The assumption that returns on stocks follow a random walk with positive drift, can be compared with $p > .5$ in our study. If, in real markets, investors under-estimate the compound probability of a stock to come up ahead after a few periods (like we show in our stylized example), they will regard the stock as riskier than it really is. That may be a partial explanation of the equity premium puzzle. The puzzle refers to the fact that over the last century the risk-return relationship has been so much more favorable for stocks than for bonds, that unreasonably high level of risk avers-ion would be needed to explain why investors are willing to hold bonds at all (Mehra and Prescott [1985]). Could it be the case that it is not the level of risk aversion that is 'wrong', rather it is a case of misjudgment of risk? [5] It may be, as Arrow [1982] argues, that evidence of bounded rationality which is found in many cases in stylized experimental work, may teach us about risk perception in complex financial markets. It could also be the case that, like is commonly argued by economists, that even when individuals make systematic mistakes markets are not biased (see e.g. Camerer [1987]). This we would like to investigate in future research.

[5] Note that this kind of irrationality does not allow for arbitrage against the irrational individual.

Appendix

Your name: ....

Welcome to our experiment in decision theory. In the experiment you will be presented with a problem, and asked to estimate the chance of a certain outcome. The more accurate your estimation will be, the more money you will earn. After you will finish answering we will pay you according to the following rule:

You will start with f 20, and for every 1% of mistake, f 1 will be deduced from your payoff. The mistake is the absolute value of [your guess (in percentages) minus the actual chance].

For example, if you will guess accurately, you will get f 20. If you will make a 10% mistake (either over-estimate or under-estimate), you will get f 10. If your mistake will be bigger than or equal to 20% you will not be paid at all.

The problem is based on the following: Mr. X is given a stock that is worth f 3 today. He will hold the stock for 10 years. In each year, the price of the stock either goes up or down by f 1. The chance of it going up is 2/3 (i.e. 67%), and the chance of it going down is 1/3 (i.e. 33%). So, after the first year, the stock will be worth either f 2 (with 33% chance) or f 4 (with 67% chance). In the second year, again, the price of the stock will either go up or down by f 1, with the same chances, and so on. If the price of the stock will reach f 0 within the 10 years, then the stock will be worthless for all future periods. Otherwise, Mr. X will sell the stock after 10 years.

The problem :

What do you think is the chance that the price of the stock after 10 years will be at least f 3? .... %

After all of you will finish answering, we will collect and check the answers, and pay you as described above.

Do you have any questions?

References

Arrow, K.J. (1982) "Risk Perception in Psychology and Economics," *Economic Inquiry*, 1-9.

Bar-Hillel, M. (1973) "On the Subjective Probability of Compound Events." *Organlzatonal Behavior and Human Performance*, 396-406.

Benartzi, S., and R. Thaler (1996) "Risk aversion or myopia? the fallacy of small numbers and its implications for retirement saving," UCLA Working Paper.

Camerer, C. (1987) "Do Biases in Probability Judgment Matter in Markets? Experimental Evidence." American *Economic Review,* 981-997.

De Bondt, W and R. Thaler (1989) "Further Evidence on Investor Overreaction and Stock Market Seasonality." *Journal of Finance*, 557-581.

Fama, E.F. (1965) "The Behavior of Stock Markets Prices." *Journal of Business,* 34-105.

Fama, E.F. (1991) "Efficient Capital Markets: II." *Journal of Finance*, 1575-1617.

Gigerenzer, G. (1996) "On narrow norms and vague heuristics: a reply to Kahneman and Tversky." *Psychological Review,* 592-596.

Gneezy, U. (1996) "Probability Judgements in Multi-Stage Problems: Experimental Evidence of Systematic Biases." *Acta Psychologica*, 59-68.

Hampel, F.R., E.M. Ronchetti, P.J. Rousseeuw, and W.A. Stahel (1986) *Robust Statistics,* Wiley, New York.

Kahneman, D. and A. Tversky (1996) "On the reality of cognitive illusions." *Psychological Review,* 582-591.

Kahneman, D., P. Slovic, and A. Tversky (1982) *Judgment Under Uncertainty: Heuristics and Biases,* New York: Cambridge University Press.

Mehra, R. and E.C. Prescott (1985) "The Equity Premium Puzzle." *Journal of Monetary Economics*, 145-61.

Stael von Holstein, C.A. (1970) *Assessment and Evaluation of Subjective Probability Distributionstions,* The Economic Research Institute at the Stockholm School of Economics.

Wagenaar, W.A. and S.D. Sagaria (1975) "Misperception of Experimental Growth." *Perception and Psychophysics,* 422-426.

Chapter 4

Strategic Delegation: An Experiment[1]

Abstract: This paper examines the effects of strategic delegation in a simple ultimatum game experiment. Our main concern is to examine the way delegation alter the way players think about the game and play it. Specifically, we show that when the proposer uses a delegate, his share *increases*. Since in such a case the proposer does not use the delegate as a commitment device, this effect identifies an additional explanation to the delegation phenomena. We also show that unobserved delegation by the responder *lowers* his share as his delegate is perceived to be more willing to accept tough offers. In such a case the results of the experiment is closer to the game theoretic prediction of the ultimatum game.

[1] This is a joint project with Chaim Fershtman.

1. Introduction:

In many types of games players, instead of playing the game themselves, prefer to send agents that play the game on their behalf. Why do players use agents to play games? There are several possible explanations of this phenomena. The first is that in some games, players choose agents who have special skills that make them better players. For example, players may send lawyers to negotiate on their behalf if the knowledge of the law may yield an advantage in the negotiation. A second possible explanation, for the delegation phenomena, is that players may send agents when they are under the impression that these agents are more intelligent or more experienced than they are and therefore may play the game better than they do. This explanation, however, relies on a bounded rationality argument in which some players are more able than others (they can either think faster, calculate all the possible contingencies, think about creative alternatives etc.), and where these abilities are important for playing the game. The third explanation is that delegation may serve as a commitment device; that is, the mere possibility of using an agent may give the player an advantage in the game as it allows him to commit to a certain behavior. The role of delegates as a commitment device has been coined in the literature as *strategic delegation* and has been extensively discussed since Schelling (1960)[2].

The main structure of a delegation game entails an additional primary stage in the game where players may hire delegates and either give them instructions on how to play the game or sign compensation scheme contracts which reward the delegates according to their performance.

---

[2] For the different aspects of strategic delegation see Caillaud, Jullien, and Picard (1995), Fershtman and Judd (1987), Fershtman , Judd, and Kalai (1991), Fershtman and Kalai (1997), Gal-Or (1996), Green (1990), and Katz (1991).

The compensation scheme may, or may not, be publicly observable. The possibility of observing the delegate's compensation scheme may drastically affect the outcome of the game. When the agent's compensation scheme is observable and irreversible, it serves as a commitment device manipulating the agent's strategic behavior and consequently the outcome of the game. The observability assumption has drawn harsh criticism in the literature. Critics have claimed that when the compensation schemes are not observable, delegation cannot serve as a commitment device (see Katz (1991)).[3] While the intuition of this claim may be convincing, the formal analysis is not obvious. In a recent paper, Fershtman and Kalai (1997) analyzed simple ultimatum games with unobserved delegation and showed the conditions under which delegation, even when it is unobservable, may affect the outcome of the game.

In this paper we examine the effects of strategic delegation in a simple ultimatum game experiment. Our main concern is to examine the effect of delegation on the way players think about the game and play it.[4] We therefore extend the discussion on delegation and consider the possibility that the use of delegates, by itself, may affect the way players perceive the game and consequently the outcome of the game.

The standard ultimatum game is a two-player game in which at the first stage, one of the players, denoted as the Proposer, proposes a division of a given pie between himself and the other player. At the second stage, the other player, denoted as the Responder, either "accepts" or "rejects" the offer. Acceptance is followed by executing the division while

---

[3] See also Dewatripoint (1988) for a discassion on the role of delegation as a commitment device when the compensation scheme can be renegotiated.

[4] The role of agency in bargaining games was considered also by Schotter, Snyder and Zheng (1995). The main issue in that paper was the effect of agency on the efficiency of the bargaining. That is, do we expect a grater breakdown of the bargaing process when it is executed by agents rather than by the original players themselves.

rejection implies that both players get no share of the pie. This type of ultimatum game has been extensively discussed in the literature (for recent surveys see Camerer and Thaler (1995), Guth (1995), and Roth (1995)). While theory implies that, at equilibrium, the Proposer gets all (or almost all) of the pie, experiments show that most divisions are not so extreme and that the average offer is typically between 30 and 50 percent, with many 50:50 splits. Moreover, low offers (20 percent or less) are frequently rejected.

Into the above ultimatum game setup, we introduce agents that represent either the Responder or the Proposer. We let the players provide compensation schemes (either observable or unobservable) for the agents and then examine how the game is played and how it differs from the ultimatum game without delegation. Thus our main focus is not on the difference between the outcome of the experiments and the equilibria identified by the theory but on the effects of the different types of delegation on the outcome of the ultimatum game experiment.

Using a messenger to deliver bad messages (or, in our case, bad offers) is a commonly observed practice. Would a Responder react identically to the same offers if made directly by the Proposer or by the Proposer's messenger or agent? This is not a simple issue. In doing ultimatum game experiments, the outcome usually differs from theoretical subgame perfect equilibrium. Arguments like a taste for fair division[5], norms of behavior, etc., are commonly used in order to explain the deviation from the theoretical predictions (again, see the surveys by Camerer and Thaler (1995), Guth (1995), and Roth (1995)). That is, the Proposer refrains from making an "unfair" offer as he is afraid that such an offer will be rejected simply on the

[5] The meaning of "fair" and "unfair" is usually exogenously given and determined by the norm of behavior in the society. It may vary across societies, groups, genders, etc..

basis of being unfair. However it is possible that the same Responder is willing to accept the *same* offer from an agent if he knows that it is not the agent who benefits from the unfair division and, moreover, that in punishing the Proposer for an unfair offer, the agent will also be punished automatically. Similarly, would an agent that represents the Responder be as sensitive as the Responder himself, to "unfair" offers? After all there is no reason for the agent to take such offers personally as it is the Responder who is treated unfairly.

Indeed our experiment indicates that the Proposers' payoffs are significantly higher when they use delegates. Note that in such a game the Proposer has the ability to make "take it or leave it" offers. Thus the advantage from using an agent is not from using it as a commitment device, but simply because the participation of the agent in the game induces a different behavior from the other player, i.e., the Responder. A possible explanation of this phenomena is that the delegate's offer is more easily accepted by the Responder as the offer is not given directly by the Proposer but by a third party. Moreover, the Responder may be less keen to punish the Proposer since by doing so he punishes also the delegate. Given such a behavior the Proposer optimally provides incentives to his agent to give tough offers

Our experiment indicates that unobserved delegation by the Responder *reduces* his share. A possible explanation of this phenomena is that the agent is perceived to be more willing to accept tough offers. That is, the willingness of the delegate to punish the Proposer for an "unfair" proposal made to a third party (the Responder) is lower than the willingness of the Responder himself to punish for a direct unfair proposal. Since the Proposer figures this effect in advance, he concludes that he can make a more greedy proposal with a lower risk of being rejected. Note that in such a case the outcome of our experiment is closer to the game theory equilibrium of the ultimatum game.

2. Setup and design of the Delegation Game

We conducted four experimental sessions, administrated in writing, and held in regular class rooms. In sessions 1,..,4 we had 60, 42, 51, 39 participants, respectively (192 in total). Participants were mostly first-year economic students recruited voluntarily in their classes. They were informed that the experiment would consist of two parts, but that they would be informed about the instructions for the second part only after completing the first.

Part I in all sessions was a simple ultimatum game. In this game, 100 'points' were to be divided between two players, a "Proposer" and a "Responder" [6]. At the first stage of the game, the Proposer proposed a division of the 100 points. If the Responder accepted the division, then both players got their shares. If the Responder rejected the offer, then both players received zero. (The instructions for part I are given in Appendix 1).

Consider now the possible use of delegates in the above ultimatum game. Delegates can be used either by the Proposer or by the Responder[7]. The delegation contract may be either observable or unobservable. Part II of the experiment (which differed across sessions) examined the following four variations of ultimatum games with delegation.

Clearly when an experiment consists of two parts there may be some degree of learning which takes place. While most of our conclusions are derived from comparing the second parts

[6] We used points instead of money in order to have a cake of 100. The conversion rate we used was 5 points = *f* 1. At the time of the experiment, September 1996, *f* 1.6=$1.

[7] The possibility exists that both the Proposer and the Responder will employ agents, but we do not consider such a case in this paper.

of the experiments, we also compare the outcome to the benchmark outcome from the first part. Here it is important to point out that Guth, Schmittberger and Schwarze (1982) studied an ultimatum game played by players who play the game twice without finding any significant difference between the two plays (see also Roth (1995)).

Delegation by the Proposer: In the first session, hereafter PO game (observable delegation by the Proposer), the Proposer uses a delegate to make the proposal on his behalf. An extra 20 points are available to the Proposer exclusively for use in providing an incentive scheme for the delegate. That is, if after delegating the action and providing the incentive scheme, not all the 20 points are paid to the delegate, none of the original players may claim the remaining points. Under such rules, delegation is costless; the pie to be divided between the Proposer and the Responder remains of the *same* size with or without delegation, which enables a simple comparison between the different scenarios that we investigate.

The procedure for Part II of the first session is as follows: At the first stage, the Proposer hires an Agent and signs a publicly observed compensation contract that specifies the Agent's fee as a function of the number of points the Proposer will receive.[8] At the second stage of the game, the Agent proposes a division of the 100 points and the Responder needs to reply by "accept" or "reject". The final division is similar to the original ultimatum game (Part I) wherein the delegate receives the points according to his compensation scheme, but only if the Responder accepts the proposal (i.e., the payoff to the Agent is also contingent on

[8] A variation of this problem would be to compensate the delegate on the basis of the proposal that he is making, independently of whether the offer is accepted or rejected.

whether the proposal is accepted or rejected). The instructions for this part are given in Appendix 2.

The second session of the experiment, hereafter PN game, is the same as the PO game but the delegate's compensation scheme in this case is *not* observed.

The subgame perfect equilibrium of the PO game is as follows: the Proposer provides the Agent with the compensation scheme of paying him 20 points (or any other positive amount) if he proposes 99 points to him and 1 point to the Responder, for any other proposal, the delegate will receive zero points. The delegate indeed offers the division 99:1 and the Responder accepts. The equilibrium of the PN game is the same as that of the PO game.[9]

Do we expect any strategic delegation in games PO and PN? According to the structure of the game itself, the Proposer has the power to make "take it or leave it" offers. In such a case, the possibility of using a delegate does not benefit the Proposer. Our first hypothesis is based on this intuition; That is, the outcome of the PO and the PN games would be the same as the outcome in the regular ultimatum game.

The competing hypothesis is that the Proposer may benefit from the use of a delegate. The rationale for such a hypothesis is that the Proposer may use the delegate as a shield that allows him to indirectly give, by means of the delegate, bad offers. That is, if the Proposer suggests a division in which he takes most of the points he runs the risk that the Responder will "reject" the proposal in order to punish him for an "unfair" offer. It is not clear that the Responder will react the same to an "unfair" offer that comes from a third party. Moreover, if the Responder rejects the offer, he punishes not only the original Proposer but also the

[9] One can also support the 100:0 division as a subgame perfect equilibrium.

"innocent bystanding" agent. That is, if we accept the view that players may choose to punish offers that are unfair, even at some cost to themselves, it is nonetheless unclear whether they are willing to punish players who are not to be blamed. In such a case, the delegate may be viewed as a *hostage*.

We do not have a specific hypothesis for the PN game as the above "hostage" argument also holds for this case. The question is, of course, if it is possible to use the agent as a hostage even when the contract with him is unobservable.

Delegation by the Responder: In the third session of the experiment, hereafter RO game, it is the Responder who is using a delegate that will respond to the offer made by the Proposer. The Responder may use the extra 20 points to provide the agent with an incentive scheme. The RO game proceeds as follows: At the outset of the game, the Responder signs a *publicly observed* contract with the delegate. At the second stage the Proposer, after observing the delegate's compensation scheme, makes his proposal of the division of the 100 points. At the last stage, the delegate either accepts or rejects the offer.

The fourth session of the experiment, hereafter RN game, is the same as the RO game but in this case the delegate's compensation scheme is *unobserved*. That is, the Responder is using an agent but the compensation scheme that he provides to this agent cannot be observed by the Proposer.

Regarding the RO game, we examine the hypothesis that the use of observable delegation, as in the PO game, affect the outcome of the game by providing an advantage for the player who hires a delegate.

In considering the role of observability we compare the outcomes of the experiment of the RO game with that of the RN game in order to examine three competing hypotheses. The first one is that delegation, when it is unobservable, is in- affective and thus the outcome of the RN game will not be significantly different from the outcome of the original ultimatum game. This hypothesis is in the spirit of Katz (1992), who argues that in the RN game, delegation does not affect the outcome of the game; in particular, the Responder cannot benefit from strategic precommitment. The (rational agent) equilibrium of this game, as suggested by Katz, is that the Responder provides the compensation scheme: "I will give you 20 points as long as you accept any positive offer". The Proposer then offers the division of 99 to himself and 1 for the Responder and the delegate "accepts" such a proposal.

The second hypothesis is that the Responder may benefit from using an agent even when the incentive scheme, provided by him, is not publicly observed. This hypothesis is in the spirit of Fershtman and Kalai (1997), who showed that commitment via delegation may be beneficial even when the compensation scheme is unobservable. The potential for such benefits depends on the type of delegation (incentive versus instructive), the possibility of repetition, and the probability of observability.

The third competing hypothesis is that the Responder will be *worse off* from using an agent. That is, once the Proposer uses an agent and the incentive scheme is unobserved, the proposals, as well as his expected payoffs, will be lower[10]. In such a case the Responder is clearly better off without using an agent.

[10] We wish to delay the rationale for such a hypothesis to our discussion section.

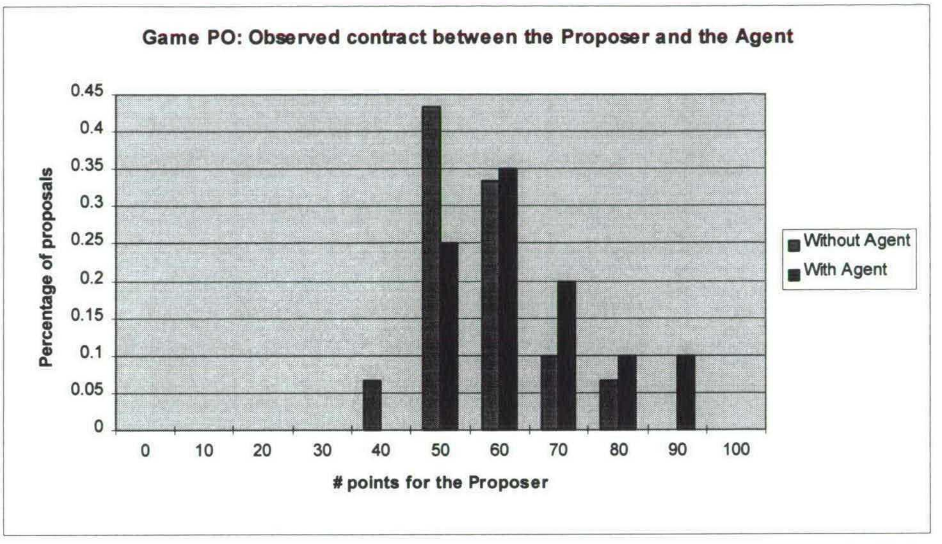
3. Results

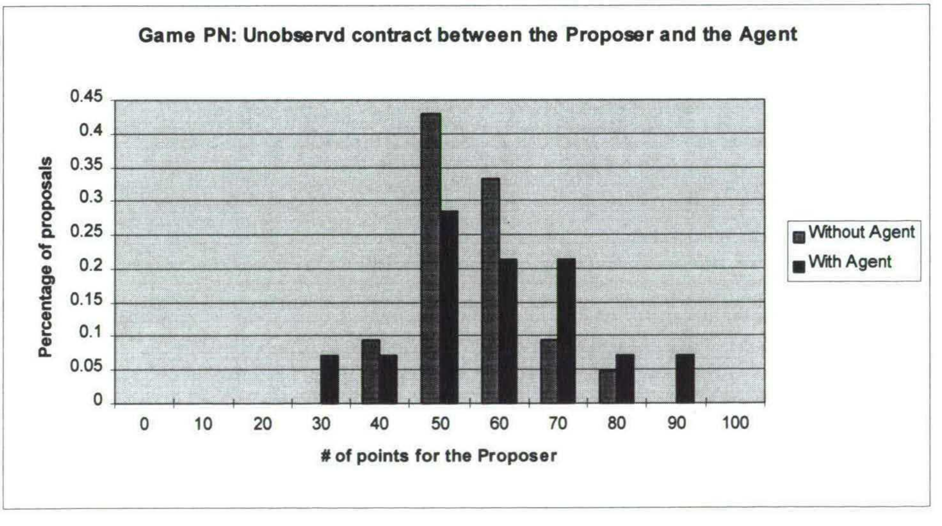
The basic question in each of the four types of delegation games, in our experiment, is whether the use of a delegate changes the outcome of the game and under what circumstances a Proposer (or Responder) may expect to benefit (or suffer) from the use of a delegate. The outcome of our experiment is described in Appendix 3, Table A1, in which we present all the proposals that were made in each of the four games, including proposals that were rejected. In Table 1, below, we present the average proposal and the average payoffs (taking into account the rejections) for each part of our four games.

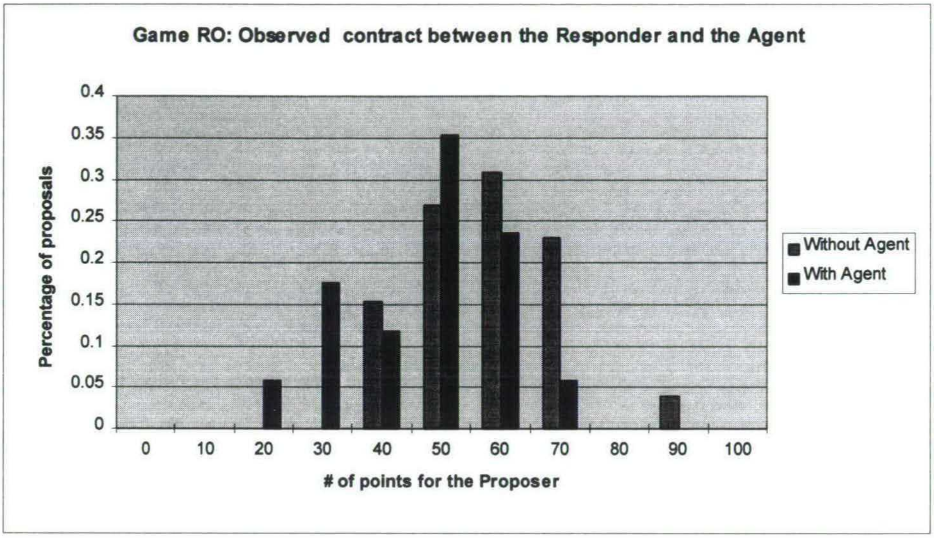
	PO Game	PN Game	RO Game	RN Game
Without Agent:				
Ave. Proposal	56.67	55.71	57.69	55.50
Ave. profit for Proposer	47.67	49.52	49.23	48.00
Ave. Profit for responder	38.96	40.96	39.23	42.00
With Agent:				
Ave. Proposal	64.50	59.29	47.06	66.92
Ave. profit for Proposer	60.50	52.86	39.41	57.69
Ave. profit for Responder	36.50	40.00	48.82	26.93

Table 1: The average proposal and the average payoffs in the four games.

In the first part of Table 1, we present the results for the first part of the experiment, in which players played the ultimatum game without delegation. In the second part of the table we present the average proposal and payoffs (to both the Proposer and the Responder) in the four delegation games that we studied. Before elaborating on these results, it would be useful to describe the distribution of the proposals that were made in each variation of the delegation game. This is done in Figure 1.


Game PO: Observed contract between the Proposer and the Agent
0.45
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
Percentage of proposals
0
10
20
30
40
50
60
70
80
90
100
points for the Proposer
Without Agent
With Agent


Game PN: Unobservd contract between the Proposer and the Agent
0.45
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
Percentage of proposals
0
10
20
30
40
50
60
70
80
90
100
# of points for the Proposer
Without Agent
With Agent



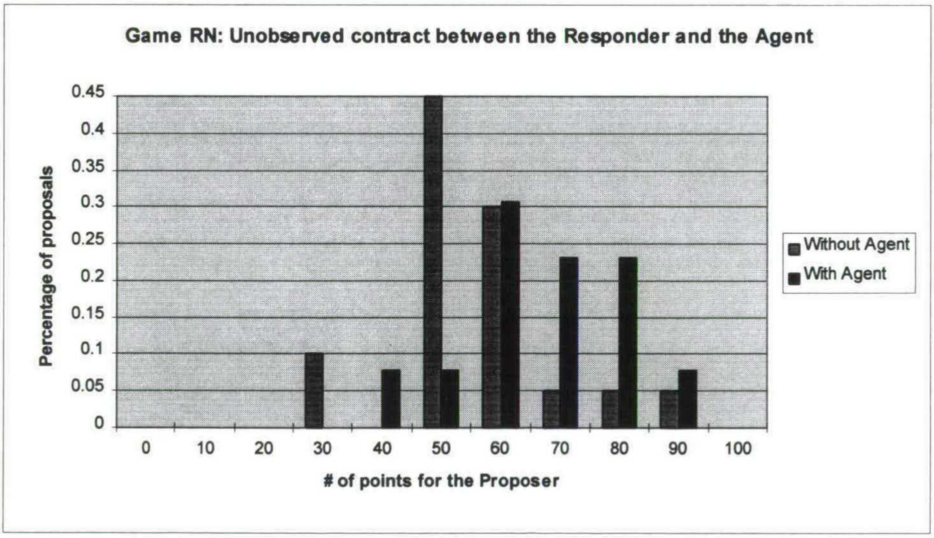


Figure 1: The distribution of proposal that were made in each variation of the delegation game.

Before turning to a more formal testing of our results, we provide a pairwise comparison of the outcomes of the ultimatum games of the different games (see Appendix 4). Our test indicates that there is no significant ex-ante difference between the groups.

Now we turn to test our hypothesis regarding the different effects of delegation. To do so we compare, for each game, the outcomes of Part I (the ultimatum game) with the outcomes of Part II (the delegated game). For comparison, we use the Mann-Whitney U test. We report the test results in Table 2.

	PO Game	PN Game	RO Game	RN Game
Profit-Proposer	.0119	.5007	.0455	.0682
Profit-Responder	.1273	.6616	.0483	.0326
Proposal	.0349	.4588	.0254	.0224

Table 2: Mann-Whitney U tests with pairwise comparisons of the medians of outcomes in Part I and Part II of each game. (The numbers are the probability of a result larger than $|z|$, where z is the test statistic).

PO Game: When the Proposer uses an agent with an observable compensation scheme, the average proposal went up from 56.7 to 64.5 and the average payoffs to the Proposer went up from 47.7 to 60.5 (see Table 1). Observing Table 2, it is evident that when using a delegate, the Proposers gave significantly (at a .95 level of significance) higher proposals (a larger share to themselves and a lower share to the Responder), and their profits were significantly higher as well.

PN Game: From Table 1 one can see that when the Proposer uses an agent but the compensation scheme is unobserved, the average proposal goes *up* from 55.7 to 59.3 while the Proposer's payoffs increases from 49.5 to 52.9. These changes are in the same direction as in the PO game but, as indicated in Table 2, these changes are not significant.

RO Game: When the Responder uses an agent and the contract is observable, the average proposal *declines* from 57.7 to 47.1, the average payoffs of the Proposer *declines* from 49.2 to 39.4 while the Responder's average payoffs increase from 39.2 to 48.8. From Tables 1 and 2 we learn that the use of a delegate by the Responder significantly improves both the proposals that he receives and his payoffs provided that the agency contract is observable.

RN Game: In the RN game, the Responder uses an agent but the agency contract is unobserved. From Tables 1 and 2 we learn that the unobserved delegation induces significant changes in the offers made and the payoffs received by both players. The average proposal *increases* from 55.5 to 66.9, the Proposer's average payoff *increases* from 48.0 to 57.7, while the Responder's average payoffs decreases from 42.0 to 26.9[11]. Surprisingly, the effect of unobserved delegation, in this case, is in the opposite direction than in the RO case, in which the agency contract is observable. Thus, the use of an agent with unobserved contract makes the Responder *worse* off.

[11] One of the two rejections in Part II of game RN is problematic. The Proposer in this observation offered a division of 60:40; the Responder offered the Agent 20 points for accepting this offer (contract 6 in Appendix 5d), yet the Agent rejected the proposal. We report on all our observations, but note that the 'spirit' of the above discussion would not change even if we did not take this observation into account.

4. Discussion: The different effects of delegation.

In the regular ultimatum game, it is the Proposer who has the power to make "take it or leave it" offers, therefore, the theory suggests that he will receive all the surplus. In such a case, there is no role for agency as a commitment device. Yet the results of our PO session indicate that the Proposers' payoffs are significantly higher when they use agents. This result implies that an additional explanation for the effectiveness of delegation exists. In the regular ultimatum game, the Proposer realizes, when making his offer, that he might be punished for making an "unfair" offer. He also understands that although the Responder is willing to punish him for an "unfair" proposal, this willingness decreases in the presence of a delegate because punishing the Proposer would imply punishing an "innocent" delegate as well. In other words, the Proposer uses the delegate as a hostage. Note that indeed in the PO session, four out of 30 (i.e. 13%) proposals were rejected in the ultimatum game, but only one out of 20 (i.e. 5%) in the game with the agent-although the overall proposals were significantly higher in the delegated game.

In the PN game we did not identify any significant effect of delegation. Casual observation of Figure 1 indicate an increase in the variance of the offers. We however prefer, not to draw any specific conclusion from this part of the experiment beyond the statement that the observability of the incentive contract changes the way players play the game.

In the RO game, it is the Responder who uses an agent. In such a case, the agent serves as a commitment device. At the first stage of the game, the Responder signs an observable compensation scheme with the agent, which allows him to commit not to accept certain offers. Our experiment indicates that the Responder benefits from the delegation and his expected payoffs increase significantly.

We find the outcome of the RN part of the experiment the most surprising. For this part, we identified initially three competing hypothesis. The first one is that RN delegation does not affect the outcome of the game. The Responder cannot use the agent as a commitment device because the incentive contract is not observable. The second hypothesis is that even without observability there is some commitment value in delegation; therefore, the Responder will benefit from the use of agents. We found out that we can reject these two hypothesis and that, to our surprise, the Responder should expect to end up *worse off* from using an agent with unobserved contract. The explanation we suggest for this result is that the willingness of the delegate to punish the Proposer for an "unfair" proposal *made to a third party* (the Responder) is lower than the willingness of the original Responder to punish for a direct unfair proposal. Moreover, the Proposer figures this effect in advance, and concludes that he can make a more greedy proposal with a lower risk of being rejected.

The above result is in contrast to Katz (1991) and Fershtman and Kalai (1997). Katz (1991) argues that the use of a delegate with an unobserved contract will not influence the outcome of the game (i.e., the outcomes will be similar to those of the ultimatum game). Fershtman and Kalai (1997) predict that, in many cases, the use of a delegate influences bargaining even if the contract is unobserved, and thus the effect of some delegations is in the direction of the RO prediction.

Note that while our experiment examines a game with unobserved delegation, it cannot be viewed as an experiment that evaluates the different claims of Katz (1991) and Fershtman and Kalai (1997). It has already been well established that the outcome of ultimatum bargaining experiments differs from the theoretical subgame perfect equilibrium of this game.

Thus, observing a difference in the outcomes of the RO and the RN games may be due to the frequently observed deviation of these experiments from the equilibrium prescribed by game theory rather than an indication of the theoretical role of unobserved delegation. To our opinion, the contribution of the experiments that compare the outcomes of the RO and RN games with the original ultimatum game without delegation is to see to what degree the use of delegation is helpful and whether players take advantage of strategic delegation even when it is unobservable.

Comparing the incentive contracts provided in the RO game and in the RN game indicates that the Responder indeed understands the role of delegation as a commitment device. In the RO game the Responders provided an "aggressive" incentive contracts. Observe that the median value for which he is giving all the 20 points to the agent is the amount of 80 to the Responder. In the unobserved case the Responder realizes that the unobservability implies that agency does not have a commitment value, and the median value for the agent to receive all the 20 points decreased to 20 (see the table in Appendix 5d).

5. Concluding Remark.

In this paper, we have described an experiment designed to analyze the effect of delegation on the outcomes of ultimatum games. The main conclusion of this experiment is that delegation significantly changes the outcome of the game. Beyond the standard explanations of strategic delegation, our experiment suggests that the introduction of an additional player, the agent in our case, changes the players' perceptions regarding the norm of behavior and what constitute a fair division in the game they are playing. These suggestions may be extended beyond the scope of ultimatum games and delegation. There are many games in which the strategic interaction may determine the entrance of a new player into the game; for example, in market games in which entry deterrence is possible and the firms' actions may affect the possibility of entrance. In such cases, changes in the set of players may affect the players' perception about the (fair) norm of behavior or other behavioral rules that the players prefer to obey. Such perceptions affect the way that these type of games are played, and therefore changing these perception should be discussed in a strategic context.

Appendix 1: The introduction and instructions for part I

Introduction

The instructions are simple, and if you follow them carefully you may earn a considerable amount of money that will be paid to you in cash at the end of the experiment. 60 students participate in this experiment. Each of you is about to get an envelope with a number. This is your registration number. Please look at it and then put it back in the envelope without letting anyone else see it. At the end of the experiment you will be asked to show the registration number you have in the envelope to the experimenter, and he will pay you according to your performance. Do not forget to write your registration number on all the forms that you will get.
The experiment consists of two parts.

Instructions for part I

In this part, 100 points are to be divided between two persons: the "Proposer" and the "Responder". At the end of the experiment, each of the two persons will get 20 cents for each point he will have.
A proposal about how to divide the 100 points between the two persons is made by the Proposer. Upon receiving the proposal the Responder is asked to respond by either accepting or rejecting it.

(a) If the Responder accepts the proposal, then both he and the Proposer are paid according to the proposal.

(b) If the Responder rejects the proposal, then both persons are paid 0 points.

The procedure for Part I is as follows: 30 students will be selected randomly to play the role of the Proposer in this part. Each Proposer will get a form on which he is asked to indicate his proposal to the Responder. The proposal must be in multiples of 10 (0, 10 , 20, 30, etc.). For example, either 0 to the responder and 100 to the proposer, or 10 to the responder and 90 to the proposer, etc.

After the Proposers will make their choice we will collect all the forms in a box, and let each of the 30 Responders students to pick randomly one form out of the box. The Responder will not be able to know what is written on the form before choosing it, and will never know the identity of the Proposer with whom he was matched (he will only know the registration number of that person). The Responder is asked to indicate on the form whether he accepts or rejects the proposal. We will collect the forms and write down the payment for each student for this part (using the registration numbers). Then part II will start. You will get the instructions for part II after part I will be over.

Appendix 2: The instructions for part II of game PO

Instructions for part II

This part is similar to part I, but this time the Proposer can not make the proposal himself. Instead, the Proposer must hire an "Agent" to make the proposal on his behalf. First, each Proposer will write a contract with an Agent. The Agent will see the contract before deciding how much to propose to the Responder. After the Agent will make the proposal the Responder will see **both** the proposal and the contract between the Proposer and the Agent. Then, the Responder will be asked to decide whether to accept or reject the proposal.
In order to pay the Agent, the Proposer gets 20 points (which he can use only to pay the Agent). If the Proposer offers the Agent less than 20 points, then the rest of the points are lost.

The procedure for Part II is as follows: 20 students will be selected randomly to play the role of the Proposer in this part. Each of them will get a form with the following table

Payment from the Proposer to the Agent

# points for the Proposer	0	10	20	30	40	50	60	70	80	90	100
# points to the Agent											

In each column the Proposer is asked to write how much to pay the Agent if he gets for him the corresponding number of points. That is, if according to the agents proposal this amount of points will be given to the Proposer. For example, in the column of 90, the Proposer is asked to write how much to pay the Agent if he gets for him 90 points, etc. After all the Proposers will fill out this table on the form, we will collect the forms in a box.

We will then select randomly 20 students out of the remaining 40 to play the role of the Agent. Each Agent will pick randomly one form out of the box, and observe the table that the Proposer he is matched with made. The Agent is now asked to make a proposal to the Responder. The forms will be collected again in the box.

Each of the remaining 20 students will be a Responder. Each will randomly pick one form out of the box and observe both the Agent's payment table and the proposal made by the Agent. Then he is asked to decide whether to accept or reject the proposal. The Responder is asked to indicate his choice on the form.

To summaries, the procedure is as follows:

The Proposer →	**The Agent** →	**The Responder**
Writes a contract with the Agent	Observes the contract and makes a proposal	Observes both the contract and the proposal and decides whether to accept or reject the proposal

Remarks:

(a) The payment from the Proposer to the Agent does not have to be in multiples of 10.

(b) If the proposal that the Agent makes is rejected, then all persons, including the Agent, get 0 points for Part II.

We will then collect all the forms, find out how much money each of you earned in Part I and Part II, and pay each of you privately. This will end the experiment. If you have any question please raise your hand and one of the experimenters will come to you.

Appendix 3: The Proposals

#	PO Game Without Agent	PO Game With Agent	PN Game Without Agent	PN Game With Agent	RO Game Without Agent	RO Game With Agent	RN Game Without Agent	RN Game With Agent
1	80*	90	80	90*	90*	70*	90*	90
2	80	90	70*	80	70*	60*	80	80
3	70*	80*	70	70	70	60	70	80
4	70	80	60*	70	70	60	60*	80
5	70	70	60	70	70	60	60	70
6	60*	70	60	60	70	50	60	70
7	60*	70	60	60	70	50	60	70
8	60	70	60	60	60*	50	60	60*
9	60	60	60	50	60	50	60	60*
10	60	60	60	50	60	50	50	60
11	60	60	50	50	60	50	50	60
12	60	60	50	50	60	40	50	50
13	60	60	50	40	60	40	50	40
14	60	60	50	30	60	30	50	
15	60	60	50		60	30	50	
16	50	50	50		50	30	50	
17	50	50	50		50	20	50	
18	50	50	50		50		50	
19	50	50	50		50		30	
20	50	50	40		50		30	
21	50		40		50			
22	50				50			
23	50				40			
24	50				40			
25	50				40			
26	50				40			
27	50							
28	50							
29	40							
30	40							
Ave. profit Proposer	47.67	60.5	49.52	52.86	49.23	39.41	48	57.69
Ave. profit Responder	38.96	36.5	40.96	40	39.23	48.82	42	26.93
Average proposal	56.67	64.5	55.71	59.29	57.69	47.06	55.5	66.92

Table A1: The Proposals made by subjects. The proposals that were rejected are with a *.

Appendix 4: Comparing the population in the four games.

We use the nonparametric Mann-Whitney *U* test based on ranks in order to test whether the samples of the outcomes come from populations having the same median. This is the appropriate test because the distributions are not normal. We report the test results in Table A2.

	Game 1 and 2	Game 1 and 3	Game 1 and 4	Game 2 and 3	Game 2 and 4	Game 3 and 4
Profit-Proposer	.8934	.6873	.9526	.8139	.8449	.6657
Profit-Responder	.7449	.8630	.9842	.6378	.7543	.8767
Proposal	.7814	.7116	.7215	.5562	.9169	.5062

Table A2: Mann-Whitney *U* tests with pairwise comparisons of the medians of outcomes in the ultimatum game by sessions. (The numbers are the probability of a result larger than $|z|$, where z is the test statistic).

From Table A2 we learn that, with a .95 level of significance (actually, even at .5 level of significance) we cannot reject the hypothesis that each of the two samples compared are from populations with the same median.

Appendix 5: The incentive contracts in the four games.

Amount for the Proposer	**0**	**10**	**20**	**30**	**40**	**50**	**60**	**70**	**80**	**90**	**100**
1	0	10	**20**	20	20	20	20	20	20	20	20
2	0	5	10	15	**20**	20	20	20	20	20	20
3	0	0	10	15	**20**	20	20	15	15	10	10
4	0	5	10	15	18	**20**	18	0	0	0	0
5	0	0	0	5	10	**20**	15	10	5	0	0
6	0	0	0	0	0	15	**20**	10	0	0	0
7	0	0	5	5	10	10	**20**	10	0	0	0
8	0	0	5	5	5	10	10	**20**	20	20	20
9	0	0	0	0	0	10	15	**20**	0	0	0
10	0	0	0	0	0	0	0	**20**	0	0	0
11	0	2	4	6	10	15	18	19	**20**	20	20
12	0	2	4	6	8	14	16	18	**20**	20	20
13	0	2	4	6	8	10	15	18	**20**	20	20
14	0	2	4	6	10	12	15	17	**20**	20	20
15	0	2	4	6	8	10	12	15	**20**	20	20
16	0	0	0	5	5	5	10	15	**20**	20	15
17	2	4	6	8	10	12	14	16	18	**20**	20
18	2	8	5	7	0	19	13	16	0	2	3
19	8	8	10	10	11	14	14	15	15	18	18
20	9	5	0	7	8	4	3	10	15	9	3

Appendix 5a: The contracts of game PO.

Amount for the Proposer	**0**	**10**	**20**	**30**	**40**	**50**	**60**	**70**	**80**	**90**	**100**
1	**20**	20	20	20	20	20	20	20	20	20	20
2	0	5	10	15	**20**	20	20	20	20	20	20
3	0	3	9	15	**20**	20	20	20	20	20	20
4	0	8	10	12	18	**20**	20	20	20	20	20
5	0	4	8	12	16	**20**	16	12	8	4	0
6	0	0	5	5	15	18	**20**	5	5	0	0
7	0	5	8	10	12	15	**20**	20	20	20	20
8	0	0	0	0	5	10	**20**	20	20	20	20
9	0	3	7	10	13	15	19	**20**	20	20	20
10	0	0	0	8	10	15	18	**20**	10	10	10
11	0	0	5	5	10	10	15	**20**	20	20	20
12	0	0	0	0	5	10	15	**20**	20	20	20
13	0	2	4	5	6	8	10	15	**20**	20	20
14	0	2	4	6	8	10	12	14	16	18	**20**

Appendix 5b: The contracts of game PN.

Amount for the Responder	0	10	20	30	40	50	60	70	80	90	100
1	5	10	15	**20**	10	5	10	15	0	10	20
2	0	5	10	15	18	**20**	20	20	20	20	20
3	0	5	10	10	15	15	**20**	20	20	20	20
4	0	0	0	0	0	0	**20**	20	20	20	20
5	0	1	1	2	2	5	5	**20**	2	1	0
6	0	0	0	0	0	0	0	**20**	20	20	20
7	0	0	0	0	0	0	0	**20**	20	20	20
8	0	0	4	6	10	10	16	18	**20**	20	20
9	0	1	4	10	13	14	15	15	**20**	20	20
10	0	0	0	0	0	5	10	15	**20**	20	20
11	0	12	15	15	16	16	18	18	19	**20**	20
12	0	2	5	10	10	15	15	15	15	**20**	20
13	0	3	4	5	11	12	14	17	18	19	**20**
14	0	2	4	6	8	10	12	14	16	18	**20**
15	0	2	4	6	8	10	12	14	16	18	**20**
16	0	15	15	15	15	15	15	15	15	15	**20**
17	1	1	1	1	1	1	2	5	10	15	**20**

Appendix 5c: The contracts of game RO.

Amount for the Proposer	**0**	**10**	**20**	**30**	**40**	**50**	**60**	**70**	**80**	**90**	**100**
1	**20**	20	20	20	20	20	20	20	20	20	20
2	**20**	20	20	20	20	20	20	20	20	20	20
3	0	**20**	20	20	20	20	20	20	20	20	20
4	0	10	**20**	20	20	20	20	20	20	20	20
5	0	10	**20**	20	20	20	20	20	20	20	20
6	0	5	**20**	20	20	20	20	20	20	20	20
7	0	0	**20**	20	20	20	20	20	20	20	20
8	0	0	**20**	20	20	20	20	20	20	20	20
9	0	0	0	0	0	**20**	0	0	0	0	0
10	0	11	12	13	14	15	16	17	18	19	**20**
11	0	2	4	16	18	19	12	14	16	18	**20**
12	0	2	4	6	8	10	12	14	16	18	**20**
13	0	2	4	6	8	10	12	14	16	18	**20**

Appendix 5d: The contracts of game RN.

References

Caillaud, B., B. Jullien and P. Picard, (1995): "Competing Vertical Structures: Precommitment and Renegotiation," *Econometrica*, 63, 621-646.

Camerer, C. and R. Thaler, (1995) "Ultimatums, Dictators and Manners," *Journal of Economics Perspectives*, 9:2, 209-219.

Dewatripont, M. (1988) "Commitment Through Renegotiation-Proof Contracts with Third Parties," *Review of Economic Studies*, LV 377-390.

Fershtman, C. and K.L. Judd, (1987) "Equilibrium Incentives in Oligopoly," *American Economic Review*, 77, 927-940.

Fershtman, C., K.L. Judd and E. Kalai, (1991) "Cooperation Through Delegation," *International Economic Review*, 32, 551-550.

Fershtman, C. and E. Kalai, (1997) "Unobserved Delegation,", *International Economic Review* (Forthcoming).

Green, J. (1990) "Strategic Use of Contracts With Third Parties", Harvard University Discussion Paper # 1502.

Gal-Or, E. (1996) "Multiprincipal Agency Relationships as Implied by Product Market Competition", mimeo, University of Pittsburgh.

Guth, W. (1995) "On Ultimatum Bargaining Experiments: A Personal Review," *Journal of Economic Behavior and Organization*, 27:3, 329-44.

Guth, W., R. Schmittberger and B. Schwarze (1982) "An Experimental Analysis of Ultimatum Bargaining" *Journal of Economic Behavior and Organization*, 3, 367-388.

Katz, M.L. (1991) "Game - Playing Agents: Unobservable Contracts as Precommitments," *The Rand Journal of Economics*, 22:3, 307-328

Roth, A. (1995) "Bargaining Experiments," In J. Kagel and A. Roth (eds), *The Handbook of Experimental Economics*, Princeton: Princeton University Press.

Schelling, T.C. (1960) *The Strategy of Conflict*, New York: Oxford University Press.

Schotter, A. B. Snyder and W. Zheng, (1995) "Bargaining Through Agents: An Experimental Study of Delegation and Commitment", Working Paper, New York University.

Chapter 5

On Competing Rewards Standards[1]

-An Experimental Study of Ultimatum Bargaining-

Abstract: In the tradition of earlier experimental studies, this paper introduces competing reward standards by letting parties bargain over the distribution of chips. The monetary equivalents of a chip for the bargaining parties can be equal (no competing rewards) or different (competing rewards). The ultimatum game is used as a tool to learn about reward standards in an asymmetric procedure. A major effect of different monetary chip equivalents is observed only when the proposer has a higher chip value. Results are compared to those reported in Kagel et al. (1996), who used a different experimental design.

[1] This is a joint project with Werner Guth.

1. Introduction

Rewards standards measure how people perceive their success when performing a certain task. In interactive situations, such reward standards usually rely on commonly accepted views on what constitutes a reward and how to measure individual rewards. In experiments, competing reward standards can be easily introduced by allowing parties to bargain over the distribution of chips whose monetary equivalent (that is, the value of a chip) varies for different individuals. (See Nydegger and Owen, 1974, for an early application.) The two competing rewards are then the **amount of chips** that an individual receives, and the **monetary earning** implied by the chips assignment.

The original motivation for using this experimental method was to test experimentally basic axioms of game theoretic concepts (see Nydegger and Owen, 1974, and Roth and Murnighan, 1982, who were mainly interested in testing the independence of bargaining results with respect to affine utility transformations as required, for instance, by Nash, 1953). Since changing the positive monetary chip value actually amounts to a positive affine utility transformation, this change does not affect the game theoretic prediction (relying on such axioms). In this research tradition, competing reward standards are a convenient experimental method to challenge the empirical validity of a certain rationality requirement.

According to the hierarchical structure of the chips earnings versus the monetary earnings, equity theory (see Homans, 1961, for an early reference) would predict equal chip assignments when the monetary value of chips for individuals are not common knowledge. On the other hand, it predicts that monetary earnings will be equalized when values are commonly known, i.e. when the superior reward standard of monetary earnings is applicable (see Guth, 1988, and 1994). This has been demonstrated most clearly by Nydegger and Owen (1974) and subsequently by Roth and Malouf (1979). See Roth (1995) for a more comprehensive survey.

Whereas the above-mentioned studies were concerned with symmetric bargaining, e.g. the demand game of Nash (1953), the experiment reported in this paper has used the extremely asymmetric ultimatum game. In the ultimatum game, player 1 (the "proposer") first proposes how to split the total amount of chips. Then player 2 (the "responder") decides whether to accept or reject this proposal. If the responder accepts, then the proposal is implemented; otherwise, both players receive nothing. For players motivated purely by monetary considerations, the game theoretic solution implies that the proposer receives almost all the money. This is not the observed outcome in experiments. The deviation is usually attributed to "fairness" considerations.

Testing fairness in asymmetric bargaining games should not be perceived as a test of equity theory, since it is not claimed that equity considerations dominate all other, e.g. strategic considerations. What we therefore try to explore experimentally is the trade-off between fairness and strategic considerations. Moreover, the structure of the ultimatum game is such that players may develop different fairness standards depending on their role. We can thus explore whether and how relative strategic advantages will influence the standard on which one relies.

Kagel, Kim, and Moser (1996) (hereafter KKM) have also used the ultimatum game as a tool to study this phenomena. Since the KKM study is closely related to the study in this paper, it will be discussed in more detail below.

We report here the results of three different treatments: In treatment (2,1), the value of a chip for player 1 was twice its value for player 2; in treatment (1,1), they had a common value; in treatment (1,2), the value of a chip for player 2 was twice its value for player 1.

In treatment (2,1), player 1 may consider an equal chip split as "fair" since it gives him a higher reward. On the other hand, the responder may consider an equal money split as "fair", and for that reason be likely to reject an equal chip split which he conceives as unfair proposals. In the regular ultimatum game, the proposer, on average, typically claims a bit more than 50% of the cake (again, see the survey by Roth, 1995). In our case, the proposers claim a bit less in terms of the chips, but a much larger share of the money. We conclude that the average proposal is more in line with the equal chip split than the money split in this case. In treatment (1,1), both the equal chip split and the equal money split coincide. Our results in this case are in line with what is usually observed. The proposers claim a bit more than 50% of the cake. In treatment (1,2), player 1 is expected to favor an equal money split to an equal chip split. However, our result does not support this. In fact, proposals are not significantly different from the proposals of treatment (1,1).

2. Experimental procedure

Before going on to elaborate on our own procedure, will first describe the KKM procedure. In the KKM study, unequal chip equivalents could favor either the proposer or the responder ($0.10 or $0.30 per chip). The total amount of chips to be allocated was 100, and only unequal chip value was tested. Furthermore, they varied the information about the monetary chip value of the other party (own-chip values were always known). Participants in the KKM experiment played the ultimatum game in the same role (proposer or responder) ten times with different partners, learning only about their own plays. One of the ten successive plays was then finally selected by chance for actual payment.

In the current study, we focused on the "full information" condition. That is, the conversion rates were commonly known. The reason is the interest in the hierarchy of reward standards. We find some of the results obtained by KKM for this condition striking:

(i) High rejection rates (39% of all proposals in the case when the proposer has the higher value).

(ii) Proposers for the most part refrained from proposing equal earnings when they had the higher value per chip.

(iii) When the proposers had the lower value per chip, their mean proposals were consistent with the equal-earnings prediction.

The rejection rates are quite high compared with other experiments (see Roth 1995). Tendencies (ii) and (iii) imply that proposers apply the superior reward standard when this is in their own advantage.[2] The current experiment was conducted to test the robustness of these result with respect to the procedure.

We had three treatments, with 100 chips to be divided in each game. In treatment (2,1), the value of a chip was 0.4 Guilders for the proposer and 0.2 Guilders for the responder. In treatment (1,1) the value of a chip was 0.2 Guilders for each player, and in treatment (1,2) it was 0.2 for the proposer and 0.4 Guilders for the responder. The value of the chips was commonly known in all treatments. The game was played only once.[3]

---

[2] Such a behavioral tendency is in contrast to the politeness ritual, observed in reward allocation experiments (Shapiro, 1975).

[3] We were interested in testing whether results (ii) and (iii) are robust for higher monetary incentives. To guarantee this, participants played only once (see footnote 4 of KKM, which acknowledges this problem). The value of the pie was about $18 in our experiment, compared with $20/10=$2 in the KKM experiment.

Compared with KKM, we have therefore used less dramatic differences in monetary chip equivalents and included a treatment with equal equivalents, which enables us to compare our results to other studies of ultimatum games. Moreover, our participants played only once in order to increase the monetary incentives.

The participants in the experiment were undergraduate students in economics at the University of Tilburg. Students were recruited in classes. Each treatment was conducted with a different group of participants. The instructions they were given are presented in Appendix A.

3. Results

The result of the plays (16, 14, and 15 in treatment (2,1), (1,1), and (1,2) respectively) are presented in Appendix B. We use the nonparametric Mann-Whitney *U*-test based on ranks to test the following two hypotheses:

1. The distribution of *chips* is not affected by the different treatments, and
2. The distribution of *money* is not affected by the different treatments.

The results are presented in Table 1.

	(2,1) and (1,1)	(2,1) and (1,2)	(1,1) and (1,2)
Chips-split	.05*	.02*	.95
Money-split	.00*	.00*	.95

Table 1: Pairwise comparisons of the distribution of chips and money for the different treatments. The numbers in the table are the *p*-values. *indicates significant differenes.

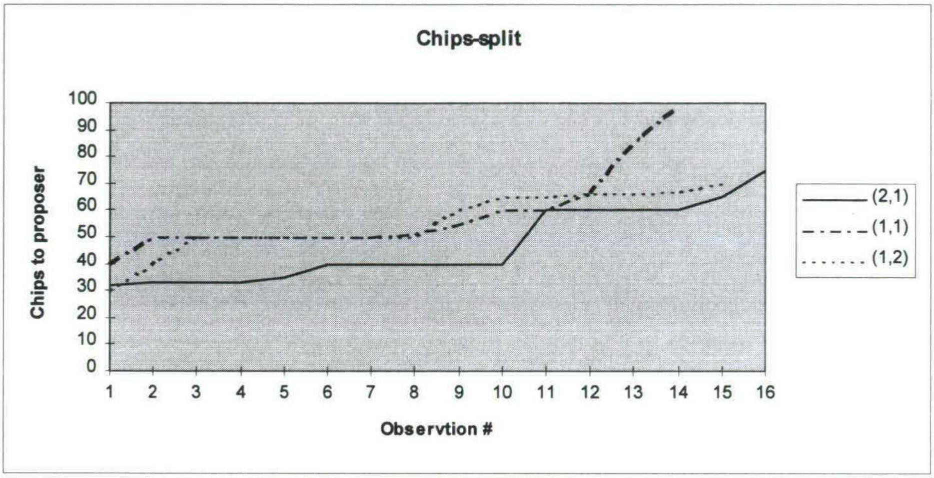
We cannot reject the hypothesis that the chip-split, as well as the money-split, in treatments (1,1) and (b1,2) comes from distributions with the same mean. Both these hypotheses are rejected, however, when we compare treatment (2,1) to the other two treatments: Proposers in treatment (2,1) asked significantly *less chips for themselves* than in the other treatments, but significantly *more money*. These comparisons are presented in Figures 1a and 1b.

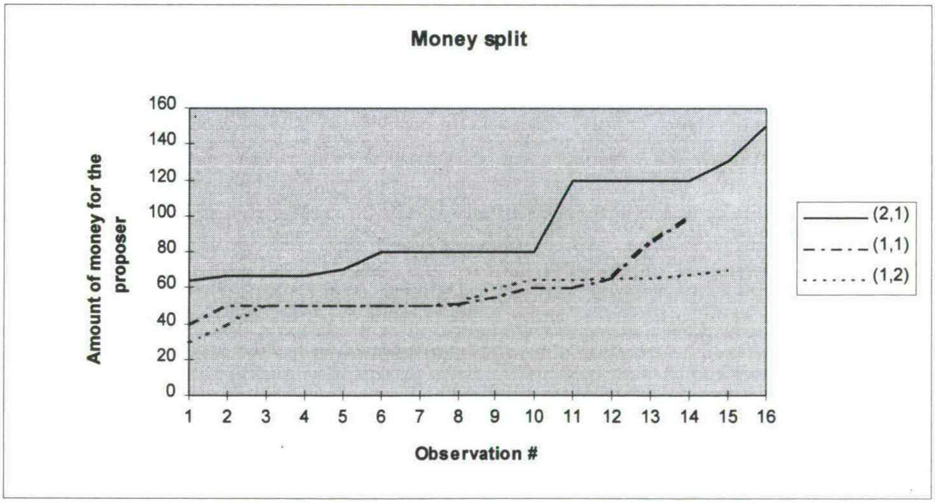
Note that equal earnings would require

-the (33, 67) or (34, 66)-chip assignment for treatment (2,1)

-the (50, 50)-chip assignment for treatment (1,1)

-the (33, 67) or (34, 66)-chip assignment for treatment (1,2).





Figures 1a and 1b: Comparisons of of chips split and money split according to treatments.

Treatment	2,1	1,1	1,2	All
Hit rate of basic reward standard (chips)	0 of 16 (0%)	7 of 14 (50%)	6 of 15 (40%)	13 of 45 (29%)
Hit rate of superior reward standard (money)	4 of 16 (25%)	7 of 14 (50%)	6 of 15 (40%)	17 of 45 (38%)
Hit rate of equity theory in general	4 of 16 (25%)	7 of 14 (50%)	12 of 15 (80%)	23 of 45 (51%)

Table 1: Hit rates of proposals in line with the basic and superior reward standard and of equity theory in general (a hit is given when the observations deviates by 5 chips or less from the prediction).

The hit rate of equal earnings is 25%, 50%, and 40% for treatment (2,1), (1,1), and (1,2), respectively. For the basic chip standard it is 0% for treatment (2,1) and 40% for treatment (1,2). Finally, only 51% of all observations can be justified by equity considerations.[4]

Remember, however, that this does not question the validity of equity theory: In the asymmetric ultimatum game equity considerations and strategic aspects are conflicting. However, it is interesting to observe whether behavior deviates from that implied by strategic aspects toward a more equitable results (as is partly true for the KKM data).

Comparing our results with those of the KKM study, we observed a dramatically lower rejection rate (overall less than 9%). The equal earning result (iii) is also rejected by our data. The only consistent observation is their tendency (ii) stating that most proposers with larger chip equivalent refrain from granting equal earnings, but try to stay close to the 50:50-chip

[4] A standard test of equity theory is not obvious, since, without allowing for any error or noise, any violation would reject it. One possibility would be to specify alternative hypotheses, e.g. the one of uniformly distributed proposals over some range, and to test their relative success. Here we do not engage in such an attempt.

distribution. Counting earning differences smaller than or equal to 5 chips as equal, 5 of the 16 proposers in the (2,1)-treatment aimed at equal earnings, as compared to 7 out of 14 in the (1,1)-treatment and 5 out of 15 in the (1,2)-treatment. Thus, it is not so much the share of proposers aiming at equal earnings which differs, but more the direction and size of the deviations.

A double ultimatum hypothesis claiming that the proposer cannot only dictate the chip allocation, but also the reward standard would have predicted the 50:50 or a nearby chip allocation in the case of (2,1) and the 67:33 allocation in the (1,2) treatment as the equitable benchmark. Whereas in the second case the predictive success of this equitable benchmark (allowing for deviations greater than or equal to 5 chips) is 40%, no 50:50 or nearby allocations has been observed for the (2,1) treatment: six proposers took considerably more and ten considerably less than 50 chips.

Another way to describe the different results for the (2,1) and the (1,2) treatment is to distinguish between three groups of participants: Those who ask for (at least ten chips) more than 50 chips, those who ask for less than 50 chips, and those who allocate the chips evenly. According to Table 2, the group of 50:50 proposers is largest for treatment (1,1), still substantial in treatment (1,2), but non-existent for treatment (2,1). Thus, the more basic chip-standard is completely ruled out when it would favor the responder: If proposers care for fair rewards, they invariably rely on the superior rewards of monetary earnings. If they do not, they try to exploit their ultimatum power by asking for even more than 50 chips.

	Proposer's demand		
Treatment	Less than 50	50	More than 50
2,1	10	0	6
1,1,	1	8	5
1,2	2	6	7

Table 2: Proposers demand freaquencies.

4. Discussion

Our results are quite different from those reported in KKM. First, we observe dramatically lower rejection rates. Second, we cannot confirm their observation that proposers aim at equal earnings when their monetary chip equivalent is smaller than the one of the responders.

What details in the experimental procedure could have caused these differences?[5] Unlike their counterparts in the KKM study, the participants in the current experiment played only once; learning effects may thus be different. However, hardly any learning effects are visible in the KKM data (see their Figure 1 on p.104).

The two aspects that we believe make the difference are, first, the less extreme asymmetry in chip equivalents, and second, the salience of monetary incentives ($18 instead of $2 per game). For example, if the responder's chip equivalent is only one third of the proposers's value, and only one out of ten games is paid, it may seem "less costly" and thus "more attractive" for the responder to reject a 50:50 chip allocation which denies the superior reward standard, than would be the case in our procedure.

[5] We would like to emphasis that KKM were interested in the role of information, which influenced their choice of procedure.

The KKM results and our observations *together* may provide a more complete picture to understand reward standards of asymmetric bargaining situations.

Appendix A: Instructions/Decision Forms

Instructions for the Proposer

Welcome to this experiment in decision making. Soon you will be randomly matched with another student. In the experiment, 100 points is to be divided between yourself and the other student. You are called the Proposer and he/she is called the Responder.

We will ask you to make a proposal about how to divide the 100 points between yourself and the Responder. Then we will ask the Responder to decide whether to "accept" or "reject" your proposal.

(a) If the Responder accepts the proposal, then each of you will earn points according to the proposal you made.

(b) If the Responder rejects the proposal, then neither of you will earn any points at all.

At the end of the experiment you (the Proposer) will receive 20 cents for each point you have. The Responder will receive 40 cents for each point he/she has. That is, he/she will receive twice the amount of money for each point held.

If you have no questions, please write down your ANR number and your proposal.

Your ANR number:_____________

Your Proposal:

# of points for the Proposer (you):	# of points for the Responder:
$\boxed{}$	$\boxed{}$

Please note that the numbers in the two boxes should add up to 100.

Instructions for the Responder

Welcome to this experiment in decision making. Soon you will be randomly matched with another student. In the experiment, 100 points is to be divided between yourself and the other student. You are called the Responder and he/she is called the Proposer.

We asked the Proposer to make a proposal about how to divide the 100 points between him/herself and you. Now we ask you to decide whether to "accept" or "reject" his/her proposal.

(a) If you accepts the proposal, then each of you will earn points according to the proposal made.

(b) If you rejects the proposal, then neither of you will earn any points at all.

The Proposer received similar instructions to yours. At the end of the experiment the Proposer will receive 20 cents for each point he/she has. You will receive 40 cents for each point you have. That is, you will receive twice the amount of money for each point held.

If you have no questions, please write down your ANR number and whether you accept or reject the proposal written below.

Your ANR number:______________

The Proposal made by the Proposer:

# of points for the Proposer: #of points for the Responder (you):

Your decision (please write accept or reject):______________

Appendix B: The Proposals

	2,1 Proposals	1,1 Proposals	1,2 Proposals
1	75*	99*	70*
2	65	85	67
3	60*	66	66
4	60	60	66
5	60	60	65
6	60	55	65
7	40	51	60
8	40	50	50
9	40	50	50
10	40	50	50
11	40	50	50
12	35	50	50
13	33.3	50	50
14	33.3	50	40
15	33.3		30
16	32		
Average	46.7	58.3	55.3

The proposals (in chips). * indicates proposals that were rejected.

References

Guth, W. (1988) "On the Behavioral Approach to Distributive Justice -- A Theoretical and Experimental Investigation," in *Applied Behavioral Economics* (ed. Shlomo Maital), Vol. II, Brighton, 703 - 717.

Guth, W. and R. Tietz (1988) "Ultimatum Bargaining for a Shrinking Cake -- An Experimental Analysis," in *Bounded Rational Behavior in Experimental Games and Markets* (eds. W. Albers, R. Selten, R. Tietz), *Lecture Notes in Economics and Mathematical Systems,* Vol. 314, Berlin, 111-128.

Guth, W. (1994) "Distributive Justice -- A Behavioral Theory and Empirical Evidence," in *Essays on Economic Psychology* (eds. H. Brandstatter and W. Guth), Springer - Verlag, 153 - 176.

Homans, G. C. (1961) *Social Behavior: Its Elementary Forms*, London.

Kagel, J. H., C. Kim, and D. Moser (1966) "Fairness in Ultimatum Games With Asymmetric Information and Asymmetric Payoffs," *Games and Economic Behavior*, Vol. 13, 100 - 110.

Nash, J. F. (1953) "Two-Person Cooperative Games," *Econometrica*, Vol. 21, 128 - 140.

Nydegger, R. V., and G. Own (1974) "Two-Person Bargaining: An Experimental Test of the Nash Axioms," *International Journal of Game Theory*, Vol. 3, 239 - 249.

Roth, A. E., and M. W. K. Malouf (1979) "Game - Theoretic Models and the Role of Information in Bargaining," *Psychological Review*, 86, 574 - 594.

Roth A.E., and Murnighan (1982) "The Role of Information in Bargaining: An Experimental Study," *Econometrica*, Vol. 50, 1123 - 1142.

Roth A. E. (1995) "Bargaining Experiments," in *The Handbook of Experimental Economics* (eds. J. Kagel and A. E. Roth), Princeton, N.J., Princeton University Press, 253 348.

Shapiro, E. G. (1975) "Effects of Future Interaction on Reward Allocation in Dyads: Equity or Equality," *Journal of Personality and Social Psychology*, Vol. 31, 873 - 880.

Chapter 6

Efficiency, Reciprocity, and Expectations in an Experimental Game[1]

Abstract: In an experimental 2-player game player 1 chooses to take x (Dutch guilders), in which case player 2 gets 0, or to leave the money. In the latter case payoffs are determined by player 2, who chooses how to split 20 between the players. The parameter x is varied between treatments, taking values of 4, 7, 10, 13, and 16 guilders.

The number of players choosing to leave the money is declining in x. When x is equal to 4,7, or 10 guilders many players 2 choose to allocate y to player 1 such that $y \geq x$. There is no positive correlation between x and y. Independently of x, the choices of player 2 resemble the choices made in an experimental Dictator game. We explicitly measure the players' beliefs. When x=7, 10, or 13 many players 1 choose to leave x despite expecting to get back less than x. There is a positive correlation between y and 2's expectation of 1's expectation of y.

[1] This is a joint project with Martin Dufwenberg.

1. Introduction

Suppose you find a wallet in the street. No one sees you. The wallet contains money, and some other stuff which is of apparent value to the owner but of no use to you. You can either keep the wallet, or bring it to a nearby police station for the owner to pick up. The police will routinely register your name, and subsequently ask the wallet owner to reimburse you in the amount he considers appropriate. What would you do?

It is commonly assumed in economics that people are motivated only by material, self-centered concerns. In the above situation such an assumption leads to an inefficient outcome. The owner will not reimburse the finder if he picks his wallet up at the police station. The finder figures this out and simply keeps the wallet. *Both* these persons would prefer that the owner gets back the wallet and reimburses the finder sufficiently.

A special instance of this situation is modeled in the game $\Gamma(x)$ in Fig. 1, where payoffs are in Dutch guilders (*ƒ*), x is an exogenously given parameter such that $0<x<20$, and y is chosen by player 2 such that $0 \leq y \leq 20$.

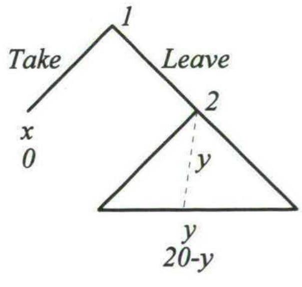


Figure 1: The game $\Gamma(x)$

If the players are motivated solely by personal monetary gain, the unique subgame perfect equilibrium in $\Gamma(x)$ is (*Take*, y=0), corresponding to the dismal outcome with the lost wallet. This outcome is inefficient since if 1 chooses *Leave* and 2 chooses y such that $x<y<20$, then a payoff vector will be realized which is better for both players.

However, much experimental evidence suggests that when humans interact they may be motivated by various non-material considerations and not only personal monetary gain. In some cases this may eliminate inefficiency. We address related issues by investigating the nature of strategic behavior in an experimental game which resembles $\Gamma(x)$. However, for methodological reasons, we ask player 2 to report a *strategy*—a choice of y—without informing her about 1's choice.[2] We still refer to the experimental game as $\Gamma(x)$.

We use x as a treatment variable taking values of *f* 4, 7, 10, 13, and 16.[3] As in $\Gamma(x)$, the money pie to be split by player 2 is held fixed at *f* 20 in all treatments. We let participants engage in anonymous, one-shot plays of this experimental game. Our objective is to record regularities in the participants decision making and to draw conclusions about the motivations behind the their behavior. To shed some additional light on this, we explicitly measure some of the players' beliefs about one another's actions and beliefs by asking the participants to make certain guesses about other participants' choices or guesses, rewarding them for accuracy.

We investigate the following issues which relate to the treatment variable x:

[2] The strategy approach, which goes back to Selten (1967), has the advantage that we can record 2's behavior irrespective of whether her information set is reached. However, 2 is not faced with the *fait accompli* of 1 choosing *Leave*, and this may affect her behavior. We take the experimental evidence reported by Schotter, Weigelt & Wilson (1994) to be indicative that this may be important (cf their findings for the games 1M, 1S, and 1H, which have similarities with $\Gamma(x)$). Since Schotter *et al* used matrices or graphs to describe their experimental games to participants while we use words only, it is not quite clear what effect to expect though. We do not explore the issue here. See Roth (1995, pp 322-3) for a discussion of the strategy method.

[3] At the time of the experiment 20 Dutch guilder was worth approximately 12 US dollars.

- An efficient outcome obtains if and only if player 1 chooses *Leave*. However, if 1 chooses *Leave* his potential loss is increasing in x. Is the propensity for 1 to choose *Leave* negatively correlated with x?

- By choosing *Leave*, player 1 places a trust in 2. To what extent does 2 reciprocate and keep this trust by choosing $y \geq x$?[4]

- One might argue that the higher is x, the kinder is 1 by choosing *Leave* since the potential loss he may incur by doing so is higher. 2 may want to be kind in return by correspondingly choosing a higher y. Is y positively correlated with x?

Furthermore, we investigate the following issues which relate to the players' beliefs:

- Is there a connection between 1's expectation of y and 1's propensity to choose *Leave*? In particular, does 1 choose *Leave* only if he expects to get back at least x?

- Is player 2's choice of y positively correlated with her expectation of 1's expectation of y (conditional on 1 choosing *Leave*)? This could happen for example if 2 were averse to "letting 1 down", in the sense of not wanting to choose y below 1's expectation of y. However, since player 2 does not know 1's expectation of y, it may be that the higher is her *expectation* of 1's expectation of y, the higher she may be inclined to choose y.

$\Gamma(x)$ is related to the *Dictator game* in which one player decides how to divide f 20 between himself and another (dummy) player. The subgame of $\Gamma(x)$ where 2 moves, considered in

[4] The usage of the terminology "place a trust", "keep the trust", and "reciprocate" here is in line with that of Berg, Dickhaut & McCabe (1995, p 126) who study a game (further discussed below) which is related to ours.

isolation, has precisely such a structure.[5] When the Dictator game is tested experimentally, with monetary payoffs controlled, "the dictator" quite often gives away more than zero, which is typically explained with reference to altruism or fairness considerations (see Davis & Holt 1993, pp 263–9 for a discussion). We suspect that 2's behavior in $\Gamma(x)$ will be affected by similar concerns, but that in addition it may matter that whether 2's choice affects payoffs is at 1's discretion. To check this, we also run an experimental Dictator game in which the procedures, including the belief elicitation, were analogous to those discussed above. Letting $y \in [0,20]$ denote the Dictator's choice of how much to allocate to the dummy player, we investigate the following issues:

- Is the choice of y lower in the Dictator game than in $\Gamma(x)$?
- Is the dictator's choice of y in the Dictator game positively correlated with her expectation of the dummy player's expectation of y?

We conclude this introduction by discussing some other related literature: Berg, Dickhaut, & McCabe (1995) analyze a "trust game" which shares many features with $\Gamma(x)$: Player 1 is given a sum of money. He chooses how much to keep and "sends" the rest to player 2. The amount sent is tripled and given to player 2 who chooses how much to "send back".[6] Bolle (1995) reports

[5] We investigate the potential importance of some non-pecuniary concerns that arise due to a choice that *precedes* a dictator subgame. One can compare this to the Ultimatum game, in which an action is added that *succeeds* a proposed dictator division: a responder gets to accept or reject the proposed split, and in the latter case, each player gets a zero payoff. See Camerer & Thaler (1995) and Güth (1995) for detailed discussions of results in experimental Ultimatum games. See Güth & van Damme (1994) for a report on an experiment on a game which incorporates essential elements of both the Dictator and the Ultimatum game.

[6] $\Gamma(x)$ can be related to Berg *et al*'s game as follows: In $\Gamma(x)$ player 1 is given a certain amount of money (x>0). He chooses to send all or nothing to player 2. The amount sent is multiplied by a factor inversely related to x, and given

experimental results involving a game which resembles $\Gamma(x)$, except that an element of chance was added. A lottery was conducted to select four out of the 64 experimental games that were played, and *only* the participants acting in these games were rewarded according to their decisions.[7] Bolle set x=½ and did not consider the effect of changing x. In both of these studies, most participants did not behave according to the subgame perfect equilibrium with only self-interested material considerations affecting payoffs.[8]

The paper is organized as follows: Section II explains the experimental procedure. Section III presents the hypotheses we test, and the experimental results. Section IV contains a discussion of our main findings. Appendices 1-3 contain the experimental instructions.

2. Experimental procedure

The participants were recruited via an ad in the weekly students' newspaper at Tilburg University and via posters on campus. These announcements invited participants to come to our offices and "sign up" for an economic experiment on decision making. We indicated that the participants' earnings would depend on these decisions, and approximately how much money was at stake.

In total we had 5 sessions of experimental $\Gamma(x)$ games and two sessions of experimental Dictator games. 12 different pairs of students interacted in each session. In the $\Gamma(x)$ games, x was

to player 2 who makes a choice on how much money to "send back" to player 1. $\Gamma(x)$ may be viewed as more general than Berg *et al*'s game in allowing other exogenously given multiplication factors than 3, and more restrictive in not allowing player 1 to send intermediate amounts.

[7] See Bolle (1990) for a discussion of whether such a set-up skews incentives relative to the case where all participants are paid.

[8] Several other authors have conducted experimental studies in which aspects of trust, reciprocity, and efficiency are key features. See e.g. Fehr, Kirchsteiger, & Riedl (1993), Fehr, Gächter & Kirchsteiger (1997), Güth, Ockenfels & Wendel (1994), van der Heijden, Nelissen, Potters, & Verbon (1997), and McKelvey & Palfrey (1992). However, these experiments are relatively less closely related to ours (various real world market institutions are mimicked, there is no "Dictator subgame", or there are more stages).

fixed within a session and was changed between sessions to f 4, 7, 10, 13, 16. For each session we had invited 13 participants to Room A, 13 participants to Room B, and 4 extra participants to a third room to cover for no-shows. After filling Rooms A and B with 13 participants (using participants from the third room room if necessary) these where given an "Introduction" (see Appendix 1). Then, they took an envelope at random. In each room, 12 envelopes contained 12 different numbers (A1,..,A12 in Room A and B1,..,B12 in Room B). These numbers were called "registration numbers". One envelope was labeled "Monitor", and determined who was the person who checked that we do not cheat. That person was paid the average of all other participants in that session. After opening the envelopes the second part of the instruction was distributed (see Appendix 2). At this point it was stressed by the experimenter that this game will be played only once.

Participants in Room A read the instruction for this part (see Appendix 2.). In the $\Gamma(x)$ sessions, they were then asked to go to the experimenter, one at a time. They got an envelope with fx in it, and then had to go behind a curtain. Over there they had to decide whether to take the money out of the envelope or not. Then to write their registration number on a note, to put this note in the envelope, and to put the envelope in a box near the experimenter.

Participants in Room B also read the instructions for this part (see Appendix 2). They were asked to write down how much they would give to their anonymous counterpart in Room A (i.e. to choose y), conditional on him/her choosing to leave the fx in the envelope in the $\Gamma(x)$ treatments. The participants' choices, sealed in envelopes, were put in a box near the experimenter.

Then part 2 started. The participants in Room A received new instructions. They were asked to guess the average y chosen by participants in Room B in part 1, and were rewarded according to accuracy (see Appendix 3). Our intention was to provide incentives for them to state their expectation of their co-player's choice of y.[9]

The participants in Room B also received new instructions (see Appendix 3). In the $\Gamma(x)$ sessions they were asked to guess the average guess of the participants in Room A who chose to leave the money in the envelope in part 1. In the Dictator game sessions they were asked to guess the average guess of the participants in Room A. Meanwhile, an experimenter and the two monitors checked and recorded the envelopes of Room A, and matched them each with an envelope from Room B (as described in the instructions). In the end, all the payoffs from part 1 and 2 were calculated and the participants were privately paid.[10]

3. Hypotheses and results

We now present the hypotheses we test and report on our experimental findings. We first discuss the experimental game $\Gamma(x)$ and then the experimental Dictator game.

A. $\Gamma(x)$

The experimental raw data concerning $\Gamma(x)$ is given in Tables 1 and 2.

[9] We want to measure 1's expectation of the y chosen by his *co-player*, but nevertheless ask 1 to guess the *average* choice of y in the whole session. We believe this creates a superior measure. Say, for example, that a participant believes that the co-player will choose y=0 with probability ½ and otherwise choose y=*f* 10. Such a participant has an expectation of *f* 5. With the incentive scheme we use he should indeed guess *f* 5. Had we asked him to guess his co-players choice he should guess either *f* 0 or *f* 10, however.

[10] Note that while the participants were anonymous to each other, the experimenter could learn each one's decision. The study of Hoffman, McCabe & Smith (1996) shows that it is then likely that dictators give away more than with subject-experimenter anonymity. Probably a similar remark applies to $\Gamma(x)$, but the importance of "social distance" concerns in general games is a matter of some controversy. See Hoffman, McCabe, Shachat & Smith (1994) and Bolton & Zwick (1995) for some partly conflicting evidence, and Roth (1995, pp 298-302) for further discussion.

	$x = 4$		$x = 7$		$x = 10$		$x = 13$		$x = 16$	
Participant	**Choice**	**Guess**	**Choice**	**Guess**	**Choice**	**Guess**	**Choice**	**Guess**	**Choice**	**Guess**
1	*L*	4	*T*	2.5	*T*	1.25	*T*	0	*T*	0
2	*L*	5	*T*	3	*T*	5	*T*	0	*T*	1
3	*L*	8	*T*	4.5	*T*	5	*T*	0	*T*	1.65
4	*L*	8	*T*	6	*T*	10	*T*	3.5	*T*	2
5	*L*	8	*T*	7.5	*L*	6	*T*	4	*T*	2.5
6	*L*	8	*T*	8	*L*	7	*T*	5.5	*T*	3
7	*L*	8	*L*	0.75	*L*	8	*T*	6.25	*T*	4
8	*L*	8.45	*L*	4	*L*	8	*T*	9	*T*	4
9	*L*	8.5	*L*	4.75	*L*	10	*L*	4	*T*	5.5
10	*L*	8.5	*L*	6	*L*	10	*L*	6	*T*	7
11	*L*	10	*L*	8	*L*	10	*L*	9	*T*	10
12	*L*	10	*L*	9	*L*	10	*L*	9	*L*	16.05
Average	12*L*/12	7.87	6*L*/12	5.33	8*L*/12	7.52	4*L*/12	4.69	1*L*/12	4.73

Table 1: Raw data on player 1 in $\Gamma(x)$. For each treatment, the first column indicates strategy choice (*T=Take, L=Leave*), and the second column indicates the guess of the average y.

	$x = 4$		$x = 7$		$x=10$		$x=13$		$x=16$	
Participant	y	**Guess**	y	**Guess**	y	**Guess**	y	**Guess**	y	**Guess**
1	0	6.5	0	4.5	0	4	0	0	0	4.5
2	4	5	0	8	0	4.5	0	7	0	5
3	4	6	0	8	1	10	0	13	0	11
4	4	8	0	9.5	5	6	1	6	2	2
5	6	7	2	7	10	7	5	5	3	5
6	10	5	2	9	10	8	7	8	4	10
7	10	8.5	7	5	10	8.5	8	3	8	8
8	10	10	8	7	10	10	8	8	10	7.5
9	10	10	9	7.5	10	10	8	8.45	10	9
10	10	10	10	8	10	10	10	7.5	10	10
11	10	10	10	8	12	5	10	8.5	10	10
12	10	10	10	9	12.5	9	16.5	7.5	12	12
Average	7.33	8.00	4.83	7.54	7.54	7.67	6.12	6.83	5.75	7.83

Table 2: Raw data on player 2 in $\Gamma(x)$. For each treatment, the first column indicates the strategy choice, and the second column indicates the guess of the average guess of y made by the players 1 who chose *Leave*.

The following two hypotheses should find support if participants behave according to the "classical solution" (subgame perfect equilibrium when each player's payoff depends only on his monetary reward):

H_0: All players 1 choose *Take*

H_1: All players 2 choose y=0

Recall that twelve different pairs of participants interacted in each treatment. Table 3 summarizes for each treatment how many participants behaved according to the classical solution (e.g: in the f4 treatment, none out of the twelve players 1 chose *Take*. In the f16 treatment, three players 2 chose y=0).

	x=4	x=7	x=10	x=13	x=16
# of *Take*	0	6	4	8	11
# of y=0	1	4	2	2	3

Table 3: Number of choices made according to the classical solution in each treatment in $\Gamma(x)$.

It is clear by inspection of the table that the hypotheses H_0 and H_1 do not find much support.

We now move towards investigating what other patterns of behavioral regularities show up in the data. We first focus on player 1, and then on player 2.

An efficient outcome results if and only if 1 chooses *Leave*. Is the propensity for player 1 to choose *Leave* related to the size of x? By inspection of Table 3 one immediately sees that the number of cases where 1 chooses *Leave* is (apart from the f7 treatment) decreasing in x.

Next we investigate whether monetary efficiency is achieved only when player 1 expects to earn at least x. In that case the following hypothesis should find support:

H_2: 1 chooses *Leave* only if 1's expectation of y is at least x.

The procedure for measuring expectations is described in Section II and Appendix 3. The relevant data are summarized in Table 4:

	x=4	x=7	x=10	x=13	x=16
# of *Leave* choices (efficent outcomes)	12	6	8	4	1
# of *Leave* choices by players who expect to get back at least x	12	2	4	0	1
Proportion of violations of H_2	0/12	4/6	4/8	4/4	0/1

Table 4: Efficiency and H_2 for each treatment in $\Gamma(x)$.

In the *f*4 treatment every player 1 chose *Leave* and in all cases the player expected to get back more than x, so H_2 was never violated. In the *f*16 treatment we have only one observation, which is in line with H_2. However, in each of the *f*7, 10, and 13 treatments H_2 is violated on four occasions.

In the remainder of this section we focus on player 2. Her choice has bearing on monetary payoffs if and only if 1 chooses *Leave*. Thereby 1 risks losing the x he could have for sure, and he gives 2 a shot at a positive payoff. To what extent does player 2 reciprocate in the sense of choosing $y \geq x$? By inspection of Table 3, one sees that this happens quite often in the *f* 4, 7, and 10 treatments, but only happens once in the other treatments.

A related aspect is that, arguably, the higher is x, the kinder is 1 by choosing *Leave* since the potential loss he may incur by doing so is higher. Player 2 may want be kind in return by correspondingly choosing a higher y. We expect an effect of this kind to motivate participants in making their choices, and therefore test the following hypothesis which we expect to be able to reject:

H_3: y and x are uncorrelated.

We use the Wilcoxon test to investigate whether the samples of y comes from populations with the same median. We do a pairwise comparison by treatments. The nonparametric Wilcoxon test is appropriate because the distributions are clearly not normal (in fact, using the skewness and kurtosis test for normality we can reject the hypothesis that y is normally distributed at a significance level of .0007). In Table 5 we report test results. A number in the intersection of a row and a column indicates, for the corresponding pair of treatments, the probability of getting at least as extreme absolute values of the test statistic as we observe, given that H_3 is true. (The last row refers to the results in the Dictator game sessions to be discussed in the next subsection.)

	x=4	**x=7**	**x=10**	**x=13**	**x=16**
x=4	-	.1124	.6033	.3408	.3865
x=7		-	.0941	.7290	.4705
x=10			-	.2253	.3123
x=13				-	.3123
x=16					-
Dictator	.2821	.4423	.1523	1.0000	.9455

Table 5: Wilcoxon tests with pairwise comparisons of medians of y by treatments in $\Gamma(x)$ (Prob > | z|, where z is the test statistic).

Table 5 conveys a result we find surprising. At the five percent level, H_3 is not rejected for any of the pairs of treatments.

Finally, we ask whether there is positive correlation between y and 2's expectation of 1's expectation of y (conditional on 1 choosing *Leave*; we henceforth suppress this qualification). This would be in line with the idea that 2 might be "averse to letting 1 down" in the sense that she does not want to give 1 less than 1 expects to get. Of course, 2 does know 1's expectation, which is why we focus on 2's expectation of this.

H_4: y is positively correlated with 2's expectation of 1's expectation of y

We first use the Spearman rank correlation coefficient (r_S) to test for the *existence* of correlation between y and 2's expectation of 1's expectation of y. We run the test for the entire 60 observations because, as shown above, the hypotheses that the choices of y in different treatments come from the same distribution can not be rejected. We find that r_S=.40, and that H_4 can not be rejected at the five percent level (in fact, H_4 can be rejected only at levels smaller than .0016). We interpret this as support for H_4. After ensuring the existence of correlation, we measure the *degree* of correlation: There is a positive correlation of .35 between y and 2's expectation of 1's expectation of y.

The connection between y and 2's expectation of 1's expectation of y is illustrated in the diagrams of Figure 2. For each treatment (including the Dictator game to be discussed in the next subsection) the choices of the participants in the player 2 position are plotted in increasing order, with the relevant second-order expectation of y plotted alongside.

B. *The Dictator game*

The experimental raw data concerning the Dictator game is given in Tables 6 and 7.

Participant	y	Guess	Participant	y	Guess
1	0	0	13	7	3
2	0	3	14	7	7
3	0	4	15	7	8
4	2	3.5	16	7	12
5	3	5	17	8	6
6	3	8	18	10	3
7	4	5	19	10	5
8	5	5	20	10	5
9	5	7.5	21	10	10
10	5	10	22	10	10
11	6	7	23	10	10
12	7	6	24	10	10
Average				6.08	6.38

Table 6: Raw data on the dictator in the Dictator game. For each participant, the first column indicates the strategy choice, and the second column indicates the guess of the dummy players' average guess of y.

Participant	Guess	Participant	Guess
1	0	13	8
2	0	14	8
3	2	15	8
4	2	16	8.3
5	2	17	10
6	3	18	10
7	3	19	10
8	3.15	20	10
9	5	21	10
10	5	22	10
11	7	23	10
12	7.5	24	15
Average			6.54

Table 7: Raw data on the dummy player in the Dictator game. For each participant, the numbers indicates the guess of the average y.

In $\Gamma(x)$ 2's subgame is reached only if 1 chooses *Leave*. If 2 is motivated by reciprocity considerations, she might choose y higher than she would as a dictator in a Dictator game (where the choice of y has payoff consequences independently of 1's behavior). Then the following hypothesis should be rejected:

H_5: The same y is chosen in the Dictator game and in $\Gamma(x)$

Refer back to Table 5. H_5 is not rejected for any value of x. We find this surprising (although perhaps less so given that H_3 was not rejected).

Finally we test the following hypothesis (motivated along the same lines as H_4):

H_6: In the Dictator game, y is positively correlated with 2's expectation of 1's expectation of y

Again, we first use the Spearman rank correlation coefficient (r_S) to test for the *existence* of correlation between y and 2's expectation of 1's expectation of y. We find that r_S=.44, and that H_6 can not be rejected at the five percent level. We interpret this as support for the hypothesis that y and 2's expectation of 1's expectation of y are correlated. After ensuring the existence of correlation, we measure the *degree* of correlation: There is a positive correlation of .51 between y and 2's expectation of 1's expectation of y.

In the last diagram of Figure 2 the dictator choices of y are plotted in increasing order, with the relevant second-order expectation of y plotted alongside.

4. Discussion

In this section we discuss our results, focusing on players 1 and 2 in turn.

A. *Results on player 1*

The higher is x, the fewer players 1 choose *Leave* (apart from in the $f7$ treatment). We find this result quite intuitive, since the potential loss that 1 may experience by choosing *Leave* is increasing in x.

It is perhaps more surprising that several players 1 choose *Leave* even when (our estimate of) their expectation of y was lower than x. Experiments in which participants chose to give up money to other participants are not new in the literature—see the discussion in the introduction about the dictator game literature, or witness many participants' behavior in the player 2 position of our game. However, as far as we know, there is little documented evidence indicating that players are willing to give up money in a way which increases monetary efficiency in situation where they expect a co-player to treat them unfavorably.

We find this result to be interesting and we note that it was made possible because we explicitly measured beliefs. While we think measuring beliefs is often a useful thing to do, a word of caution is in order. A risk avert individual in the player 1 position may have an incentive to understate his expectation of y in order to cover himself in case 2 gives back less than he expects. This observation suggests that perhaps our finding that several players 1 choose *Leave* despite not expecting to get back x should be interpreted cautiously. More generally, it also suggests that one should be aware that it is a delicate matter how to measure beliefs and provide proper incentives.

B. *Results on player 2*

We have reported that player 2 quite often reciprocates in the sense of choosing $y \geq x$ in the *f* 4, 7, and 10 treatments, while this almost never happens in the *f* 13 and 16 treatments. This result seems consistent with the findings by Berg *et al* (1995) that in their experimental game (cf. footnote X above) many players 2 send back no less than their counterpart sent them.[11] One possible explanation of our results could be that player 2 is reluctant to give 1 more than one half of the *f* 20 that may be split. For $x \leq f10$, reciprocity in the sense that $y \geq x$ can then be achieved while maintaining the no-more-than-one-half constraint. Since reciprocity is mutual, players 2 are more likely to reciprocate at x less than or equal to 10, since 2 will assume that 1 understands the constraint. This explains the bifurcation of the data at x=10.

We find no correlation between x and y in the experimental data. Indeed, the behavior of player 2 looks much like in a Dictator game. This result may be compared to the finding of Berg *et al* (1993) that there appears to be no correlation between the amount sent and the amount sent back.[12] Also van der Heijden *et al* (1997) report a similar result. Arguably, the more money is sent, the more kind is player 1. Our set-up is different in that player 1 can only be kind in one way (by choosing *Leave*), but, in a sense, we control for how kind 1 is by using x as a treatment variable. This difference between the designs turns out to be unimportant.

We find that y is positively correlated with 2's expectation of 1's expectation of y. Such correlation is in line with the idea that 2 may be averse to letting 1 down in the sense that 2 wishes not to give 1 less than 1 expects to get. Since 2 does not know 1's expectation of y she

---

[11] Lack of reciprocity when $x \in \{13,16\}$ cannot be taken as evidence against this similarity, because in Berg *et al*'s game the "multiplication factor" (cf footnote 5) is always 3 and hence never as low as 20/13 or 20/16.

[12] In Berg *et al*'s "social history" treatment (in which participant were informed about the choices made in earlier sessions before making choices) they find "an increase in the correlation between amounts sent and payback decisions" (p 135), which suggests that this result is sensitive to the social setting.

sould judge this by her expectation of 1's expectation of y, which then will be correlated with 2's choice y. We note that such effects can be modeled by incorporating beliefs directly into a player's utility function along the lines suggested by Geanakoplos, Pearce & Stacchetti (1989).[13]

[13] Geanakoplos *et al* provide several examples of how their theory can be used to incorporate "emotions" in strategic analysis. These effects are qualitatively different from many other ideas that have been advanced to rationalize experimental data, like warm glow of giving in Andreoni (1990), altruism in Andreoni & Miller (1994), aversion to unfair treatment in Bolton (1991), and empathy and gratitude in Falk & Stark (1996). In these cases, the relevant utilities can be defined on strategy profiles only. By contrast, in Geanakoplos *et al*'s theory each player's utility is defined on a richer domain, which includes the player's subjective beliefs. Rabin (1993) makes use of such ideas to develop a theory of fairness.

Appendix 1

{When the participants arrived they were directed to their seats. The participants in Room A received the following written instruction. The instruction in Room B was identical except that "Room A" was substituted for "Room B" everywhere in the text, and vice versa.}

Instruction for persons in Room A

You are about to participate in an experimental study of decision making. The experiment will last about an hour. In the experiment, each of you will be paired with a different person who is in another room. You will not be told who this person is either during or after the experiment. This is Room A, the other person is in Room B. As you notice, there are other people in the same room with you who are also participating in this experiment. You will not be paired with any of these people.

After reading this instruction, we ask you to draw one envelope from this box. In the envelope you will find a note with your 'registration number', which will be used throughout the experiment. After observing this note, please put it back in the envelope so no one else will see it. You will be asked to show this note later on when you will be paid. One envelope is an exception to this rule. Instead of a number, this envelope contains the announcement 'Monitor A'. The monitor will watch us while we carry out the experiment and assist us from time to time. An analogous procedure to determine the 'registration number' and to select 'Monitor B' is used in Room B. Every student will get *f* 8 as a show up fee, and in addition you may earn money in the experiment. Some of the money will be given to you during the experiment, and the rest at the end of it. The monitor will receive a payment equal to the average payoff of all other students in the experiment. All the money will be payed in cash.

From the moment you have drawn an envelope you are no longer allowed to talk or communicate with the other participants. If you have a question, please raise your hand and one of us will come to your table. As soon as everyone has taken his/her envelope, we will distribute further instructions.

Are there any questions about what has been said up till now?

Appendix 2

{*After the participants had read the instruction they received upon their arrival and clarifying questions had be answered (these were rare), we distributed the following instruction (identical in both rooms) in the session with the treatment in which x=f4 . Substitute "f 7, 10, 13, 16" for "f4" to get the instruction participants received in the other* $\Gamma(x)$ *sessions. In the two Dictator game sessions subjects received instructions which described that game, but were otherwise analogously formulated.*}

The Procedure

The decision procedure will be as follows: Each person in Room A will get an additional *f* 4 and have two options:

(a) to take the *f* 4. In this case (s)he gives back an empty envelope, and the person with whom he/she is matched in Room B does not get to split any money.

(b) to leave the *f* 4 in the envelope. In that case the person in Room B with whom he/she is matched with will get to split *f*20 between the two of them. That is, the person in Room B decides how much of *f*20 to give to the person in Room A, and how much of it to keep.

The remainder of these instructions will explain exactly how this experiment is run: Each person in Room A will get an envelope with *f* 4 and a note, and then, one at a time, will go behind a curtain. Over there (s)he will be asked to write his/her registration number on the note and put the note back into the envelope. Then, (s)he will have to decide whether to "take it or leave it". That is, whether to "take" (and keep) the *f* 4 and give back the envelope without the money, or to "leave" the *f* 4 in the envelope. The person in Room A will be asked to put the envelope in a box near the experimenter. If the person in Room A decides to take the money, then the person with whom (s)he is matched in Room B will not get any money to split. If the person in Room A decides to leave the money in the envelope, then the person with whom (s)he is matched in Room B will get *f*20 to split between the two of them.

If the person in Room A leaves the *f* 4, then *f*20 will be made available to split between the two paired players. The split will be determined by the person in Room B. Each person in Room B will be asked to decide how much money out of *f*20 to give to the person in Room A with whom he/she is matched. The persons in Room B are asked to write their decisions on a sheet of paper which is given to them, and then to put this sheet of paper in their envelope, and the envelope in a box near the experimenter. Note that this decision by the person in Room B will be relevant only if the person in Room A chose to leave the *f* 4.

Then, Monitor A will take the box from Room A, and Monitor B will take the box from Room B. Together with an experimenter, they will match each envelope of Room A with the envelope of the person in Room B that has the same registration number, i.e. A1 will be matched with B1, A2 with B2 etc. If the envelope of the person in Room A will be empty, then no additional money will be given. If the envelope of Room A will contain the ƒ4, then the note in the envelope from Room B will determine how to split the ƒ20 between the two persons. The experimenter (with the monitors observing) will record the payoff of each of you. You will be paid at the end of the experiment.

The experiment is structured so that, apart from the experimenter, no one will know the decisions of people in either Room A or Room B. Since your decision is private, we ask that you do not tell anyone your decision either during or after the experiment.

Appendix 3

{*After the participants' choices had been collected, in each treatment they received instructions as follows.*}

Question

{*To participants in Room A only:*}

Now we ask you to guess what was the average amount that persons in Room B chose to give back to the persons in Room A. Your reward will depend on your accuracy.

{*To participants in Room B only:*}

We asked the persons in Room A to guess how much the person in Room B chose to give back to them. We now ask you to guess what was the average of the guesses of the persons in Room A, but we consider only the persons that also chose to leave the money in the envelope. In other words, we do not consider the the guesses of those who chose not to leave the money. If no one in Room A chose to leave the money, then you will be paid ƒ5 regardless of your choice. Otherwise, your reward will depend on your accuracy.

{*For all participants the instruction continued as follows:*}

In order to check whether your guess is accurate, one of the experimenters will calculate this average, from the envelopes of the persons in Room B. You will be rewarded in the following way: You will start with ƒ5, and for every 1 cent of mistake, 1 cent will be deducted from this ƒ5. The mistake is the absolute value of (your guess - the actual average). For example, if you will guess accurately, you will get ƒ5. If you miss by, say ƒ2, (i.e. your guess is either two guilders too high or two guilders too low), you will be paid ƒ3. If your mistake will be larger than or equal to ƒ5, then you will not be payed at all for this part.

Please write your guess and your registration number on this sheet, and wait for the experimenter to collect the sheets.

References

Andreoni, J. (1990) "Impure Altruism and Donations to Public Goods: A Theory of Warm Glow Giving", *Economic Journal* 100, 464-77.

Andreoni, J. and J. Miller (1994) "Giving According to GARP: An Experimental Study of Rationality and Altruism", Paper presented at the 4th Amsterdam Workshop on Experimental Economics.

Berg, J., J. Dickhaut and K. McCabe (1995) "Trust, Reciprocity, and Social History", *Games and Economic Behavior* 10, 122-42.

Bolle, F. (1990) "High Reward Experiments without High Expenditure for the Experimenter?", *Journal of Economic Psychology* 11, 157-67.

Bolle, F. (1995) "Rewarding Trust: An Experimental Study", Diskussionspapier 15/95, Europa-Universität Viadrina.

Bolton, G. (1991) "A Comparative Model of Bargaining: Theory and Evidence", *American Economic Review* 81:5, 1096-1136.

Bolton, G. and R. Zwick (1995) "Anonymity versus Punishment in Bargaining", *Games and Economic Behavior* 10, 95-121.

Camerer, C. and R. Thaler (1995) "Ultimatums, Dictators, and Manners" *Journal of Economic Perspectives* 9:2, 209-219.

Davis, D. and C. Holt (1993) *Experimental Economics*, Princeton University Press.

Fehr, E., S. Gächter and G. Kirchsteiger (1997) "Reciprocity as a Contract Enforcement Device: Experimental Evidence", Forthcoming in *Econometrica.*

Fehr, E., G. Kirchsteiger and E. Riedl (1993) "Does Fairness Prevent Market Clearing? An Experimental Investigation", *Quarterly Journal of Economics* 108:2, 437–60.

Geanakoplos, J., D. Pearce and E. Stacchetti (1989) "Psychological Games and Sequential Rationality", *Games and Economic Behavior* 1, 60–79.

Güth, W. (1995) "On Ultimatum Bargaining Experiments: A Personal Review", *Journal of Economic Behavior and Organization* 27:3, 329-44.

Güth, W. and E. van Damme (1994) "Information, Strategic Behavior and Fairness—An Experimental Study", CentER Discussion Paper #9465, Tilburg University.

Güth, W., P. Ockenfels and M. Wendel (1994) Efficiency by Trust in Fairness? Multiperiod Ultimatum Bargaining Experiments with an Increasing Cake", *International Journal of Game Theory* 22, 51-73.

van der Heijden, E., J. Nelissen, J. Potters and H. Verbon (1997) "Gift Exchange Experiments: The role of Reciprocity and the Matching Structure", Forthcoming in the *Journal of Economic Psychology.*

Hoffman, E., K. McCabe, K. Shachat and V. Smith (1994) "Preferences, Property Rights, and Anonymity in Bargaining", *Games and Economic Behavior* 7, 346-380.

Hoffman, E., K. McCabe and V. Smith (1996) "Social Distance and Other-Regarding Behavior in Dictator Games", *American Economic Review* 86:3, 653–660.

McKelvey, R. and T. Palfrey (1992) "An Experimental Study of the Centipede Game", *Econometrica* 60, 803-36.

Rabin, M. (1993) "Incorporating Fairness into Game Theory and Economics", *American Economic Review* 83:5, 1281–1302.

Roth, A. (1995) "Bargaining Experiments", in J. Kagel & A. Roth (eds.), *The Handbook of Experimental Economics*, Princeton University Press.

Selten, R. (1967) "Die Strategiemethode zur Erforschung des eingeschränkt rationalen Verhaltens im rahmen eines Oligopolexperiments", in H. Sauermann (ed.), *Beiträge zur experimentellen Wirtschaftsforschung*, Tübingen, 136-168.

Schotter, A., K. Weigelt and C. Wilson (1994) "A Laboratory Investigation of Multiperson Rationaliy and Presentation Effects", *Games and Economic Behavior* 6, 445-68.

Stark, O., and I. Falk (1996) "On the Transfer Value of Gratitude", Mimeo, Department of Economics, University of Oslo.

Center for Economic Research, Tilburg University, The Netherlands
Dissertation Series

No.	Author	Title
1	P.J.J. Herings	Static and Dynamic Aspects of General Disequilibrium Theory; ISBN 90 5668 001 3
2*	Erwin van der Krabben	Urban Dynamics: A Real Estate Perspective - An institutional analysis of the production of the built environment; ISBN 90 5170 390 2
3	Arjan Lejour	Integrating or Desintegrating Welfare States? - a qualitative study to the consequences of economic integration on social insurance; ISBN 90 5668 003 x
4	Bas J.M. Werker	Statistical Methods in Financial Econometrics; ISBN 90 5668 002 1
5	Rudy Douven	Policy Coordination and Convergence in the EU; ISBN 90 5668 004 8
6	Arie J.T.M. Weeren	Coordination in Hierarchical Control; ISBN 90 5668 006 4
7	Herbert Hamers	Sequencing and Delivery Situations: a Game Theoretic Approach; ISBN 90 5668 005 6
8	Annemarie ter Veer	Strategic Decision Making in Politics; ISBN 90 5668 007 2
9	Zaifu Yang	Simplicial Fixed Point Algorithms and Applications; ISBN 90 5668 008 0
10	William Verkooijen	Neural Networks in Economic Modelling - An Empirical Study; ISBN 90 5668 010 2
11	Henny Romijn	Acquisition of Technological Capability in Small Firms in Developing Countries; ISBN 90 5668 009 9
12	W.B. van den Hout	The Power-Series Algorithm - A Numerical Approach to Markov Processes; ISBN 90 5668 011 0
13	Paul W.J. de Bijl	Essays in Industrial Organization and Management Strategy; ISBN 90 5668 012 9

* Copies can be ordered from Thesis Publishers, P.O. Box 14791, 1001 LG Amsterdam, The Netherlands, phone + 31 20 6255429; fax: +31 20 6203395; e-mail: thesis@thesis.aps.nl

No.	Author	Title
14	Martijn van de Ven	Intergenerational Redistribution in Representative Democracies; ISBN 90 5668 013 7
15	Eline van der Heijden	Altruism, Fairness and Public Pensions: An Investigation of Survey and Experimental Data; ISBN 90 5668 014 5
16	H.M. Webers	Competition in Spatial Location Models; ISBN 90 5668 015 3
17	Jan Bouckaert	Essays in Competition with Product Differentiation and Bargaining in Markets; ISBN 90 5668 016 1
18	Zafar Iqbal	Three-Gap Analysis of Structural Adjustment in Pakistan; ISBN 90 5668 017 x
19	Jimmy Miller	A Treatise on Labour: A Matching-Model Analysis of Labour-Market Programmes; ISBN 90 5668 018 8
20	Edwin van Dam	Graphs with Few Eigenvalues - An interplay between combinatorics and algebra; ISBN 90 5668 019 6
21	Henk Oosterhout	Takeover Barriers: the good, the bad, and the ugly; ISBN 90 5668 020 x
22	Jan Lemmen	Financial Integration in the European Union: Measurement and Determination; ISBN 90 5668 021 8
23	Chris van Raalte	Market Formation and Market Selection; ISBN 90 5668 022 6
24	Bas van Aarle	Essays on Monetary and Fiscal Policy Interaction: Applications to EMU and Eastern Europe; ISBN 90 5668 023 4
25	Francis Y. Kumah	Common Stochastic Trends and Policy Shocks in the Open Economy:Empirical Essays in International Finance and Monetary Policy; ISBN 90 5668 024 2
26	Erik Canton	Economic Growth and Business Cycles; ISBN 90 5668 025 0
27	Jeroen Hoppenbrouwers	Conceptual Modelling and the Lexicon; ISBN 90 5668 027 7
28	Paul Smit	Numerical Analysis of Eigenvalue Algorithms Based on Subspace Iterations; ISBN 90 5668 026 9

29 Uri Gneezy Essays in Behavioral Economics; ISBN 90 5668 028 5

URI GNEEZY graduated in Economics at Tel-Aviv University in 1993. He carried out his Ph.D. research at the CentER for Economic Research at Tilburg University. His main research interest is behavioral decision making and bounded rationality.

Neoclassical economics is based on a model of a rational decision maker who maximizes his utility. However, a growing body of empirical evidence show that the rational decision making model fails to describe how real people behave. The question economists face is whether the empirical facts should be allowed to spoil the good story. I think they should.

A relatively new area of research in economics, which can be called "behavioral economics", is aimed at closing that gap by improving the descriptive power of models. While normative models are typically based on a set of "rational" assumptions, the descriptive models are based on assumptions which are motivated by observed behavior.

The first step in building better descriptive models is finding the relevant behavioral regularities. For that purpose economists adopted a tool, which psychologists have been using for a very long time: *Experiments*.

This thesis contains a collection of papers which form my first steps into this world.

ISBN 90-5668-028-5

CentER *Dissertation*